Error Estimation of An Ensemble Statistical Seasonal Precipitation Prediction Model

Samuel S.P. Shen and William K.M. Lau

Climate and Radiation Branch, NASA/Goddard Space Flight Center, Greenbelt, MD 20771

Kyu-Myong Kim

Science Systems and Applications, Inc., Lanham, MD 20706

Guilong Li

Department of Mathematical Sciences, University of Alberta, Edmonton, Canada T6G 2G1

August 15, 2001

Corresponding author address: W.K.M. Lau, Climate and Radiation Branch, NASA/Goddard Space Flight Center, Greenbelt, MD 20771

E-mail: lau@climate.gsfc.nasa.gov

Technical Memorandum Abstract

Error Estimation of An Ensemble Statistical Seasonal Precipitation Prediction Model

Samuel S. P. Shen[1], William K. M. Lau[2] and Kyu-Myong Kim[3]

This report describes an optimal ensemble forecasting model for seasonal precipitation and its error estimation. Each individual forecast is based on the canonical correlation analysis (CCA) in the spectral spaces whose bases are empirical orthogonal functions (EOF). The optimal weights in the ensemble forecasting crucially depend the mean square error of each individual forecast. An estimate of the mean square error of a CCA prediction is made also using the spectral method. The error is decomposed onto EOFs of the predictand and decreases linearly according to the correlation between the predictor and predictand. Since new CCA scheme is derived for continuous fields of predictor and predictand, an area-factor is automatically included. Thus our model is an improvement of the spectral CCA scheme of Barnett and Preisendorfer (1987). The improvements include (i) the use of area-factor, (ii) the estimation of prediction error, and (iii) the optimal ensemble of multiple forecasts. The new CCA model is applied to the seasonal forecasting of the United States precipitation field. The predictor is the sea surface temperature.

[1]National Research Council
[2]NASA/Goddard Space Flight Center
[3]Science Systems and Applications, Inc.

Abstract

This NASA Technical Memorandum describes an optimal ensemble canonical correlation forecasting model for seasonal precipitation. Each individual forecast is based on the canonical correlation analysis (CCA) in the spectral spaces whose bases are empirical orthogonal functions (EOF). The optimal weights in the ensemble forecasting crucially depend on the mean square error of each individual forecast. An estimate of the mean square error of a CCA prediction is made also using the spectral method. The error is decomposed onto EOFs of the predictand and decreases linearly according to the correlation between the predictor and predictand. Since new CCA scheme is derived for continuous fields of predictor and predictand, an area-factor is automatically included. Thus our model is an improvement of the spectral CCA scheme of Barnett and Preisendorfer (1987). The improvements include (i) the use of area-factor, (ii) the estimation of prediction error, and (iii) the optimal ensemble of multiple forecasts. The new CCA model is applied to the seasonal forecasting of the United States (US) precipitation field. The predictor is the sea surface temperature (SST). The US Climate Prediction Center's reconstructed SST (1951-2000) is used as the predictor's historical data. The US National Center for Environmental Prediction's optimally interpolated precipitation (1951-2000) is used as the predictand's historical data. Our forecast experiments show that the new ensemble canonical correlation scheme renders a reasonable forecasting skill. For example, when using September-October-November SST to predict the next season December-January-February precipitation , the spatial pattern correlation between the observed and predicted are positive in 46 years among the 50 years of experiments. The positive correlations are close to or greater than 0.4 in 29 years, which indicates excellent performance of the forecasting model. The forecasting skill can be further enhanced when several predictors are used.

Contents

LIST OF FIGURES

LIST OF SYMBOLS

1. $X(\mathbf{x},t)$: predictor field

2. $Y(\mathbf{y},t')$: predictand field

3. $\sqrt{A_j}$ and $\sqrt{B_j}$: area-factors, where A_j is the area of the grid box $\mathbf{x}_j$, and B_j is the area of the grid box $\mathbf{y}_j$

4. $U(t) = \int_{\Omega_X} d\Omega_X \; X(\mathbf{x},t)u(\mathbf{x})$: canonical variable for the predictor

5. $V(t) = \int_{\Omega_Y} d\Omega_Y \; Y(\mathbf{y},t)v(\mathbf{y})$: canonical variable for the predictand

6. $\rho = \mathrm{corr}(U,V)$: canonical correlation coefficient

7. $\lambda_n^X, \psi_n(\mathbf{x})$: eigenvalues and EOFs of the predictor

8. $\lambda_n^Y, \phi_n(\mathbf{y})$: eigenvalues and EOFs of the predictand

9. EOF expansions

$$
\begin{aligned}
X(\mathbf{x},t) &= \sum_n X_n(t)\psi_n(\mathbf{x})\sqrt{\lambda_n^X}, \\
Y(\mathbf{y},t) &= \sum_m Y_m(t)\phi_m(\mathbf{y})\sqrt{\lambda_m^Y}, \\
u(\mathbf{x}) &= \sum_n u_n\psi_n(\mathbf{x})/\sqrt{\lambda_n^X} \quad \text{(canonical pattern for predictor)}, \\
v(\mathbf{y}) &= \sum_m v_m\phi_m(\mathbf{y})/\sqrt{\lambda_m^Y} \quad \text{(canonical pattern for predictand)}.
\end{aligned}
$$

10. $\tilde{\Sigma}_{XY}(m,n) = [\langle X_n Y_m \rangle]$: cross-covariance matrix in spectral space

11. Discrete covariance matrices:

$$
\begin{aligned}
\Sigma_{XX}^A &= \left[\sqrt{A_i}\langle X_i X_j\rangle\sqrt{A_j}\right], \\
\Sigma_{YY}^A &= \left[\sqrt{B_i}\langle Y_i Y_j\rangle\sqrt{B_j}\right], \\
\Sigma_{XY}^A &= \left[\sqrt{A_i}\langle X_i Y_j\rangle\sqrt{B_j}\right].
\end{aligned}
$$

12. $\mathbf{e}_n^X(i) = \psi_n(\mathbf{x}_i)\sqrt{A_i}, \quad \mathbf{e}_m^Y(j) = \phi_m(\mathbf{y}_j)\sqrt{B_j}$: discrete EOFs

13. $\hat{Y}(\mathbf{y},t') = \sum_{n=1}^p \left(\sum_{k=1}^p v_n^{(k)}\rho_k U^{(k)}(t)\right)\sqrt{\lambda_n^Y}\phi_n(\mathbf{y})$: prediction formula

14. $\epsilon^2(\mathbf{y}) = \langle\left(Y(\mathbf{y},t') - \hat{Y}(\mathbf{y},t')\right)^2\rangle$: definition of the forecasting error field

15. $\epsilon^2(\mathbf{y}) = \epsilon_q^2(\mathbf{y}) + \epsilon_R^2(\mathbf{y})$: prediction error plus residual error

16. $\epsilon_m^2 = \lambda_m^Y \sum_{k=1}^{q}(1-\rho_k^2)\left(v_m^{(k)}\right)^2$: forecasting error of the mth mode

17. $\hat{Y}(\mathbf{y},t) = \sum_{n=1}^{\infty}\sqrt{\lambda_n^Y}\hat{Y}_n(t)\phi_n(\mathbf{y})$ and $\hat{Y}_n(t) = \sum_{h=1}^{H} w_n^{(h)}\hat{Y}_n^{(h)}(t)$: ensemble forecasting

18. $w_n^{(h)} = \frac{1/(\epsilon_n^{(h)})^2}{\sum_{l=1}^{H} 1/(\epsilon_n^{(l)})^2}$: optimal weight for nth mode of the predictand and the hth predictor

19. $E^2(\mathbf{y}) = \sum_n E_n^2\phi_n^2(\mathbf{y})$: ensemble forecasting error formula

20. $E_n^2 = \left(\sum_{h=1}^{H} 1/(\epsilon_n^{(h)})^2\right)^{-1}$: ensemble forecasting error for the nth mode.

1 INTRODUCTION

The great economic significance of long-term forecasting, ranging from monthly to annual, provides enormous opportunities for the forecasting research. However, long-term precipitation forecast is extremely difficult to reach satisfactory level of accuracy. Numerous forecasting methods have been developed to improve the forecasting, yet few have systematic estimate of the forecasting product's quality. Both statistical models and dynamical general circulation models (GCM) have been explored. Despite the great potential of the coupled GCM, its forecasting ability is still limited due to numerous unknown mechanisms, including land-surface processes and cloud structure. The chaotic nature of nonlinear dynamics requires both spatially and temporally dense observations to limit the computing drift when solving a stack of nonlinear differential equations in a GCM. It is not conclusive yet whether the statistical or dynamical model is the best for forecasting. At the US Climate Prediction Center (CPC), it was stated in 1998 that "neither the dynamical nor the statistical models, as a group, perform significantly better than the other" (Barnston and He, 1998). CCA, OCN (optimal climate normals), and CA (constructive analogue) are the official statistical forecasting models used in CPC. These statistical results are compared with the coupled GCM model forecasts. A subjective determination is made when announcing the NOAA's official long term forecasting.

This NASA Technical Memorandum proposes a systematic method for error estimate of statistical forecasting by using a spectral approach with empirical orthogonal functions (EOFs) basis (Shen et al., 1998). The paper also proposes an ensemble statistical forecasting scheme, which is an ensemble canonical correlation analysis. Our canonical correlation analysis (CCA) is developed for two spatially continuous fields and is a slight improvement of the scheme of Barnett and Preisendorfer (1987), referred to as BP-CCA. Our error is assessed by the mean square error field and the estimate is applied to the continuous canonical correlation analysis, referred to as C-CCA. Since our error assessment is an *a priori* estimate, it provides an estimate on the forecasting quality at the same time when the forecasting is made. Thus, the error assessment is different from the traditional cross validation approach, which is a *posteriori* approach and assesses the error after the occurrence of the forecasting event.

The error field has a spatial structure, large over the places of low forecasting skill and small over the places of high forecasting skill. Thus, the estimated error indicates the spatial

inhomogeneity of the forecasting difficulty and assesses the quality of forecasting. Another use of the error is for optimal ensemble forecasting. It is common knowledge that when making optimization related to a stochastic process, the error information has to be known. Thus, in addition to the *a priori* assessment of the forecast quality, the error is crucial when optimally combining several forecasts. In a linear ensemble forecasting, the weight of a particular forecast in the ensemble is inversely proportional to this forecast's mean square error.

Our inclusion of area-factor, optimal ensemble of several forecasts, and error estimate constitute a new development of the statistical forecasting method, which has not only added more features to the BP-CCA but also improved the forecasting accuracy.

The classical CCA algorithm requires an inversion of cross-covariance and auto-covariance matrices (see Appendix A). In the climate forecasting practice, all these matrices are not full rank because of the short history of climate observations. The BP-CCA scheme in the EOF space avoided the inversion of matrices of non-full rank. BP-CCA was a milestone work and made it possible to use CCA as an operational forecasting tool. According to Kim and North (1998, 1999), CCA is an accurate method compared with other statistical methods, such as the methods of POP (principal oscillation pattern) (von Storch and Zwiers, 1999), Markov (Xue, 2000), EOF extrapolation (Kim and North, 1998), and MCA (maximum covariance analysis) (Lau and Wu, 1999). CPC and Canadian Climate Center are using the BP-CCA for long term forecasting of temperature and precipitation. Therefore, it is important that the method be improved in a timely fashion, if it is possible. Our inclusion of the area-factor can reduce the noise of the high latitude data and hence is expected to lead to better accuracy. Further, because of our error assessment formula, the forecasting quality at a given location is assessed simultaneously as the forecast is issued. These error estimates are crucial to optimally combine the forecasts from different models and data to form an optimal ensemble forecast.

US seasonal precipitation is a difficult predictand because of its large variance. However, the clear impact of El Nino Southern Oscillation (ENSO) on the US recitation indicates that SST does have potential predictability to the US precipitation. If so, our improved CCA, which is supposed to be more accurate than the existing statistical tools, should catch a good part of the precipitation signal. Our experiments validated the above assumption.

In our forecasting experiments, the predictor is the sea surface temperature (SST). The US Climate Prediction Center's reconstructed SST (1951-2000) is used as the predictor's his-

torical data. The US National Center for Environmental Prediction's optimally interpolated precipitation (1951-2000) is used as the predictand's historical data. Our forecast experiments show that the new ensemble canonical correlation scheme renders a reasonable forecasting skill. For example, when using September-October-November (SON) SST to predict the January-February-March precipitation of the next year, the spatial pattern correlation between the observed and predicted are positive in 46 years among the 50 years of experiments. The positive correlations are close to or greater than 0.4 in 29 years, which indicates excellent performance of the forecasting model. The forecasting skill can be further enhanced when careful selection of ocean basins is made and more predictors are used.

The context of this technical memorandum is arranged as follows. Section 2 illustrates the mathematics for the C-CCA. The discretization of the C-CCA is described in Section 3, where the symmetry of the covariance matrix with an area-factor is emphasized. The prediction as a multi-variate regression of the canonical variables is presented in Section 4. Section 5 studies the prediction error estimation and Section 6 shows optimal procedures of combining several different forecasts. The characteristics of the US precipitation are contained in Section 7. The US seasonal forecasting results are included in Section 8, and Section 9 gives summary and conclusions.

2 C-CCA: A CANONICAL CORRELATION ANALYSIS BETWEEN TWO CONTINUOUS CLIMATE FIELDS

The CCA forecasting scheme is a regression between two canonical variables $U(t)$ and $V(t')$, which are weighted integrals of two corresponding variables, predictor and predictand. Here t and $t' = t + \Delta t$ signify time, and Δt is the lead time of forecast for the predictand. The higher the absolute value of the correlation between the two random variables, the more significant the regression, and consequently, the stronger linear relationship between the two random variables. To make the regression statistically significant, the absolute value of the correlation between the two canonical variables is maximized by choosing optimal weight functions.

Let random variables $X(\mathbf{x}, t)$ and $Y(\mathbf{y}, t)$ represent the predictor and predictand fields, respectively. Here $\mathbf{x}$ is the position vector defined in the predictor domain Ω_X, $\mathbf{y}$ is the position vector in the predictand domain Ω_Y, and t signifies time. The corresponding weight

functions are $u(\mathbf{x})$ for X and $v(\mathbf{y})$ for Y, which are called canonical correlation patterns and are to be determined by an optimization. Two weighted integrals are defined by

$$U(t) = \int_{\Omega_X} d\Omega_X \; X(\mathbf{x}, t) u(\mathbf{x}) \; , \tag{1}$$

$$V(t) = \int_{\Omega_Y} d\Omega_Y \; Y(\mathbf{y}, t) v(\mathbf{y}) \; . \tag{2}$$

The unit of the weight function $u(\mathbf{x})$ is [X-unit]$^{-1}$ [area]$^{-1}$ and that of $v(\mathbf{y})$ is [Y-unit]$^{-1}$ [area]$^{-1}$. Thus, the canonical variables $U(t)$ and $V(t)$ are dimensionless. Since the theory to be developed is still a stationary theory, the processed data of X and Y are often the standardized and detrended anomalies. The area is standardized by R^2, where R is the radius of Earth.

Weight functions often imply certain physical meanings of $U(t)$ and $V(t)$. For instance, if

$$u(\mathbf{x}) = \frac{1}{\text{Area of } \Omega_X},$$

and if X is dimensionless, such as the standardized anomaly, then $U(t)$ is the spatial average of the X field over the domain Ω_X. If X is not standardized, a quantity, such as the average of the spatial standard deviation, should be included in the denominator of the above formula, in order to make $U(t)$ dimensionless.

The CCA scheme searches for the optimal weight functions $u(\mathbf{x})$ and $v(\mathbf{y})$ so that the correlation between U and V reaches extrema. The corresponding weight functions $u(\mathbf{x})$ and $v(\mathbf{y})$ are called the canonical correlation patterns. The extrema can be either positive or negative. As will be seen later, the canonical patterns will be determined by an eigenvalue problem. The orthonormal eigenvectors are unique up to a positive or negative sign. Therefore, one can select the eigenvectors so that the correlations are always non-negative, and the correlation's extrema can be regarded as maxima. The details of the eigenvector selection is given later.

The correlation between U and V is denoted by ρ. Hence, it is our intention to find the optimal weight functions $u(\mathbf{x})$ and $v(\mathbf{y})$ such that

$$\rho = \max \text{Corr}(U, V). \tag{3}$$

Since the weight functions are to be determined, one can always adjust them so that U and V have unit variance. The above correlation optimization problem is then equivalent to a covariance optimization problem

$$\rho = \max \text{Cov}(U, V) \tag{4}$$

under a constraint:

$$\mathrm{Var}(U) = \mathrm{Var}(V) = 1. \tag{5}$$

This optimization problem can be solved by the method of Lagrange multiplier and the unknown functions u and v can be determined by integral equations. The Lagrange function can be written as

$$J[u,v] = \langle UV \rangle + \zeta \left[\langle U^2 \rangle - 1 \right] + \eta \left[\langle V^2 \rangle - 1 \right]. \tag{6}$$

Here, ζ and η are Lagrange multipliers, which are dimensionless and have a simple relationship with ρ (cf., eqs. (12) and (13)), and $\langle \cdot \rangle$ denotes ensemble average.

The Lagrange function can be written into another form

$$\begin{aligned} J[u,v] = \quad & \int_{\Omega_X} d\Omega_X \int_{\Omega_Y} d\Omega_Y \, \Sigma_{XY}(\mathbf{x},\mathbf{y}) u(\mathbf{x}) v(\mathbf{y}) \\ & + \zeta \left[\left(\int_{\Omega_X} d\Omega_X \int_{\Omega_{X'}} d\Omega_{X'} \, \Sigma_{XX}(\mathbf{x},\mathbf{x}') u(\mathbf{x}) u(\mathbf{x}') \right) - 1 \right] \\ & + \eta \left[\left(\int_{\Omega_Y} d\Omega_Y \int_{\Omega_{Y'}} d\Omega_{Y'} \, \Sigma_{YY}(\mathbf{y},\mathbf{y}') v(\mathbf{y}) v(\mathbf{y}') \right) - 1 \right], \end{aligned}$$

where

$$\Sigma_{XX}(\mathbf{x},\mathbf{x}') = \langle X(\mathbf{x},t) X(\mathbf{x}',t) \rangle, \tag{7}$$

$$\Sigma_{YY}(\mathbf{y},\mathbf{y}') = \langle Y(\mathbf{y},t) Y(\mathbf{y}',t) \rangle, \tag{8}$$

are covariance functions and

$$\Sigma_{XY}(\mathbf{x},\mathbf{y}) = \langle X(\mathbf{x},t) Y(\mathbf{y},t) \rangle \tag{9}$$

is the cross-covariance function between the X and Y fields.

The optimization conditions are

$$\frac{\delta J}{\delta u} = 0$$

and

$$\frac{\delta J}{\delta v} = 0,$$

which lead to the following integral equations for u and v:

$$\int_{\Omega_Y} d\Omega_Y \, \Sigma_{XY}(\mathbf{x},\mathbf{y}) v(\mathbf{y}) + 2\zeta \int_{\Omega_{X'}} d\Omega_{X'} \, \Sigma_{XX}(\mathbf{x},\mathbf{x}') u(\mathbf{x}') = 0, \tag{10}$$

$$\int_{\Omega_X} d\Omega_X \, \Sigma_{XY}(\mathbf{x},\mathbf{y}) u(\mathbf{x}) + 2\eta \int_{\Omega_{Y'}} d\Omega_{Y'} \, \Sigma_{YY}(\mathbf{y},\mathbf{y}') v(\mathbf{y}') = 0. \tag{11}$$

In the above, the integrations in each equation leave a free independent variable. The first equation's free independent variable is $\mathbf{x}$ in the predictor domain Ω_X and the second equation's $\mathbf{y}$ in the predictand domain Ω_Y. These two equations form an eigenvalue problem for the eigenfunction $(u(\mathbf{x}), v(\mathbf{y}))$ and eigenvalue $4\zeta\eta$. This eigenvalue is related to the correlation ρ between U and V. It will be shown that the relationship is

$$\sqrt{4\zeta\eta} = \rho. \tag{12}$$

More specifically,

$$\zeta = \eta = -\rho/2. \tag{13}$$

Linear integral equations are usually solved by integral transforms or series expansions. Here, the series expansion with EOFs basis is used to solve the above integral equations (10)-(11).

EOFs are defined for X-field and Y-field, respectively, and they are the eigenfunctions of auto-covariance functions,

$$\int_{\Omega_{X'}} d\Omega_{X'}\, \Sigma_{XX}(\mathbf{x}, \mathbf{x}')\psi_n(\mathbf{x}') = \lambda_n^X \psi_n(\mathbf{x}),$$
$$\int_{\Omega_{Y'}} d\Omega_{Y'}\, \Sigma_{YY}(\mathbf{y}, \mathbf{y}')\phi_n(\mathbf{y}') = \lambda_n^Y \phi_n(\mathbf{y}).$$

The EOFs are orthonormal and form a complete basis for the vector space of the field. Namely,

$$\int_{\Omega_X} d\Omega_X\ \psi_m(\mathbf{x})\psi_n(\mathbf{x}) = \delta_{mn},$$
$$\sum_{n=1}^{\infty} \psi_n(\mathbf{x})\psi_n(\mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}'),$$

where δ_{mn} is the Kroneker delta that is equal to one when $m = n$ and zero otherwise, and $\delta(\mathbf{x} - \mathbf{x}')$ is the Dirac delta defined as a generalized function. For any continuous function $f(\mathbf{x})$ over Ω_X, it has the property that

$$\int_{\Omega_X} d\Omega_X\ \delta(\mathbf{x} - \mathbf{x}')f(\mathbf{x})\ = f(\mathbf{x}').$$

According to the above relations, the unit of X's EOF is [area]$^{-1/2}$, and the unit of λ_n^X is ([X unit]^2 [area]). The EOFs for the predictand have the same orthogonality and completeness properties and similar units.

The field and weight functions are expanded in terms of EOFs:

$$X(\mathbf{x},t) = \sum_n X_n(t)\psi_n(\mathbf{x})\sqrt{\lambda_n^X}, \tag{14}$$

$$Y(\mathbf{y},t) = \sum_m Y_m(t)\phi_m(\mathbf{y})\sqrt{\lambda_m^Y}, \tag{15}$$

$$u(\mathbf{x}) = \sum_n u_n\psi_n(\mathbf{x})/\sqrt{\lambda_n^X}, \tag{16}$$

$$v(\mathbf{y}) = \sum_m v_m\phi_m(\mathbf{y})/\sqrt{\lambda_m^Y}. \tag{17}$$

In the above, $X_n(t), Y_n(t), u_n$, and $v_n(\mathbf{y})$ are dimensionless, but the unit of $u(\mathbf{x})$ is ([X unit]$^{-1}$ [area]$^{-1}$) and the unit of $v(\mathbf{y})$ is ([Y unit]$^{-1}$ [area]$^{-1}$).

The cross-covariance function can also be expressed in terms of EOFs

$$\Sigma_{XY}(\mathbf{x},\mathbf{y}) = \sum_{m,n}\langle X_n Y_m\rangle\sqrt{\lambda_n^X}\sqrt{\lambda_m^Y}\psi_n(\mathbf{x})\phi_m(\mathbf{y}). \tag{18}$$

The weighted integrals U and V defined by (1) and (2) now become

$$U = \sum_n u_n X_n, \quad V = \sum_m v_m Y_m. \tag{19}$$

The above summations are infinite series, indices m and n running from 1 to ∞. If one truncates m upto q and n upto p, then an optimization problem of $\max\rho$ in a finite-dimensional spectral space can be obtained.

In the spectral space, the correlation between U and V and its constraints can be expressed as

$$\rho = \max\left(\tilde{\mathbf{u}}'\tilde{\Sigma}_{XY}\tilde{\mathbf{v}}\right), \tag{20}$$

$$\tilde{\mathbf{u}}'\tilde{\mathbf{u}} = 1, \qquad \tilde{\mathbf{v}}'\tilde{\mathbf{v}} = 1, \tag{21}$$

because

$$\langle X_m X_n\rangle = \delta_{mn}, \qquad \langle Y_m Y_n\rangle = \delta_{mn}.$$

In the above, the cross-covariance matrix in the spectral space is

$$\tilde{\Sigma}_{XY}(m,n) = [\langle X_n Y_m\rangle], \tag{22}$$

and

$$\tilde{\mathbf{u}}' = (u_1, u_2, \cdots, u_p), \qquad \tilde{\mathbf{v}}' = (v_1, v_2, \cdots, v_q). \tag{23}$$

The superscript "$'$" of a matrix indicates the transpose of the matrix. So, $\tilde{\mathbf{u}}'\tilde{\mathbf{v}}$ is the dot product of two vectors. The dimensions of the cross-covariance matrix are the truncation order p of the EOF expansion of X by the truncation order q of Y.

In the spectral space, the eigenvalue problems (10) and (11) become

$$\tilde{\Sigma}_{XY}\tilde{\mathbf{v}} = -2\zeta\tilde{\mathbf{u}}, \tag{24}$$

$$\tilde{\Sigma}_{YX}\tilde{\mathbf{u}} = -2\eta\tilde{\mathbf{v}}. \tag{25}$$

Substituting one into the other, one obtains an eigenvalue problem of a symmetric matrix,

$$\tilde{\Sigma}_{XY}\tilde{\Sigma}_{YX}\tilde{\mathbf{u}} = \lambda\tilde{\mathbf{u}}, \tag{26}$$

or equivalently

$$\tilde{\Sigma}_{YX}\tilde{\Sigma}_{XY}\tilde{\mathbf{v}} = \lambda\tilde{\mathbf{v}}. \tag{27}$$

In general, the eigenvalues of $B'B$ are non-negative for any matrix B. Our eigenvalue λ is specifically equal to the square of the correlation, i.e., $\lambda = \rho^2$. The proof of this is simple. Multiplying (24) by $\tilde{\mathbf{u}}'$ and (25) by $\tilde{\mathbf{v}}'$, then multiplying the two resulting equations, and finally using (20) and (21), one can deduce that $4\zeta\eta = \lambda = \rho^2$. In addition, one can have

$$\rho = -2\zeta = -2\eta, \quad \tilde{\Sigma}_{XY}\tilde{\mathbf{v}} = \rho\tilde{\mathbf{u}}, \quad \tilde{\Sigma}_{YX}\tilde{\mathbf{u}} = \rho\tilde{\mathbf{v}}. \tag{28}$$

The correlation between two fields reveals the phase differences of the two fields. When positive anomalies of U and V appear at the same time, then the correlation is one. According to the definition of the correlation ρ, the value of ρ must be between -1 and $+1$. The correlation-eigenvalue relationship $\rho^2 = \lambda$ has two solutions $\rho = \pm\lambda$. Since both $\tilde{\mathbf{u}}$ and $-\tilde{\mathbf{u}}$ are both eigenvectors of the eigenvalue problem (26), one can choose the sign of $\tilde{\mathbf{u}}$ so that $\rho = \langle U(t)V(t)\rangle \geq 0$. Next section will show how to choose the correct sign.

3 DISCRETE CCA WITH AREA-FACTORS

The above continuous theory can only become computationally possible when the covariance functions Σ_{XX}, Σ_{YY} and Σ_{XY}, or $\tilde{\Sigma}_{XY}$ are known. However, in almost all the climatological applications, they are unknowns. A discrete version of CCA must be considered. Here we derive the discrete CCA from the integral version described in the previous section, hence the area-factor is automatically included. The inclusion of the area-factor is important when the

data are defined over the uniform latitude-longitude grid boxes and when the domains Ω_X and/or Ω_Y contain high latitude regions, since the data over the low latitude boxes represent larger areas and are, thus, weighted higher. The details of the covariance matrices and EOFs with area-factor are described in Shen et al. (1998).

The exact eigenvalue problem is

$$\int_\Omega \rho(\mathbf{x}, \mathbf{x}')\psi_k(\mathbf{x}')\, d\Omega' = \lambda_k \psi_k(\mathbf{x}). \tag{29}$$

Here $\psi_k(\mathbf{x})$ is the kth EOF (or mode) and λ_k is the variance (eigenvalue) of $X(\mathbf{x})$ on the kth mode ($k = 1, 2, \cdots$). The approximate eigenvalues of the above continuum eigen-problem can be estimated by a discretization procedure given by

$$\sum_j \hat{\rho}_{ij}\hat{v}_j^{(k)} = \hat{\lambda}_k \hat{v}_i^{(k)}, \tag{30}$$

where

$$[\hat{\rho}_{ij}] = \left[\sqrt{A_i}\ \rho(\mathbf{x}_i, bx_j)\sqrt{A_j}\ \right] \tag{31}$$

is the covariance matrix with area-factor, and

$$\hat{v}_j^{(k)} = \psi_k(\mathbf{x}_j)\sqrt{A_j}, \tag{32}$$

are the eigenvectors with area-factor and A_j is the area associated with the grid $\mathbf{x}_j$. For uniform latitude-longitude grid boxes, one has

$$A_j = \cos\phi_j \Delta\theta \Delta\phi \tag{33}$$

where ϕ_j is the latitude of $\mathbf{x}_j$ and the $\Delta\theta$ and $\Delta\phi$ are the zonal and meridional box dimensions respectively, which are measured in radians. The linear spatial unit (i.e., the length unit) is in the radius of earth: $R = 6376$ km.

With the above preparation, if the area of the region associated with $\mathbf{x}_i$ is denoted by A_i and that associated with $\mathbf{y}_j$ by B_j, the covariance matrices with area-factor are

$$\Sigma_{XX}^A = \left[\sqrt{A_i}\langle X_i X_j\rangle\sqrt{A_j}\right], \tag{34}$$

$$\Sigma_{YY}^A = \left[\sqrt{B_i}\langle Y_i Y_j\rangle\sqrt{B_j}\right], \tag{35}$$

$$\Sigma_{XY}^A = \left[\sqrt{A_i}\langle X_i Y_j\rangle\sqrt{B_j}\right]. \tag{36}$$

The discrete EOFs are $\mathbf{e}_n^X$ and $\mathbf{e}_m^Y$. They are related to the continuous EOFs by

$$\mathbf{e}_n^X(i) = \psi_n(\mathbf{x}_i)\sqrt{A_i}, \quad \mathbf{e}_m^Y(j) = \phi_m(\mathbf{y}_j)\sqrt{B_j}. \tag{37}$$

Here $\mathbf{e}_n^X(i)$ is the ith element of the vector $\mathbf{e}_n^X$. The magnitude of $\mathbf{e}_n^X$ is equal to one, since

$$\sum_{i=1}^{J_X} \left(\mathbf{e}_n^X(i)\right)^2 = \sum_{i=1}^{J_X} \left(\psi_n(\mathbf{x}_i)\sqrt{A_i}\right)^2 \approx \int_{\Omega_X} d\Omega_X \; \psi_n^2(\mathbf{x}) = 1,$$

where J_X is the total number of grid points in domain Ω_X.

In practice, the discrete EOFs are computed from the eigenvalue problems of the covariance matrices with area factors.

$$\Sigma_{XX}^A \mathbf{e}_n^X = \lambda_n^X \mathbf{e}_n^X,$$
$$\Sigma_{YY}^A \mathbf{e}_n^Y = \lambda_n^Y \mathbf{e}_n^Y.$$

The correlation ρ defined by (3) has a discrete version

$$\rho = \mathrm{Corr}(\mathbf{X}_A \cdot \mathbf{u}_A, \mathbf{Y}_B \cdot \mathbf{v}_B), \tag{38}$$

where the vectors are defined by

$$\mathbf{X}_A(t) = \left[\sqrt{A_i} X(\mathbf{x}_i, t)\right], \tag{39}$$
$$\mathbf{Y}_B(t) = \left[\sqrt{B_i} Y(\mathbf{y}_i, t)\right], \tag{40}$$
$$\mathbf{u}_A = \left[\sqrt{A_i} u(\mathbf{x}_i)\right], \tag{41}$$
$$\mathbf{v}_B = \left[\sqrt{B_i} v(\mathbf{y}_i)\right]. \tag{42}$$

The maximization of the correlation ρ can then be realized by maximizing the covariance

$$\rho = \mathbf{u}_A' \Sigma_{XY}^A \mathbf{v}_B, \tag{43}$$

under the constraints

$$\mathbf{u}_A' \Sigma_{XX}^A \mathbf{u}_A = 1, \tag{44}$$
$$\mathbf{v}_B' \Sigma_{YY}^B \mathbf{v}_B = 1. \tag{45}$$

In order to find the maximal correlation ρ by solving a single eigenvalue problem, both the fields and weight functions are expressed in terms of discrete EOF. The EOF expansions are

$$\mathbf{X}_A(t) \approx \sum_{n=1}^{p} X_n(t) \mathbf{e}_n^X \sqrt{\lambda_n^X}, \tag{46}$$
$$\mathbf{Y}_B(t) \approx \sum_{m=1}^{q} Y_m(t) \mathbf{e}_m^Y \sqrt{\lambda_m^Y}, \tag{47}$$
$$\mathbf{u}_A \approx \sum_{n=1}^{p} u_n \mathbf{e}_n^X / \sqrt{\lambda_n^X}, \tag{48}$$
$$\mathbf{v}_B \approx \sum_{m=1}^{q} v_m \mathbf{e}_m^Y / \sqrt{\lambda_m^Y}. \tag{49}$$

The discussion about the truncation orders p and q is given in the appendix.

The requirement

$$\frac{1}{K_X}\sum_{t=1}^{K_X} X_m(t)X_n(t) = \delta_{mn}, \qquad \frac{1}{K_Y}\sum_{t=1}^{K_Y} Y_m(t)Y_n(t) = \delta_{mn}$$

forces the lengths of the data streams for predictor and predicatand to be the same, i.e., $K_X = K_Y = K$. The covariance matrices between the data point i and the data point j are computed by

$$\Sigma^A_{XX}(i,j) = \frac{1}{K}\sum_{t=1}^{K} X(\mathbf{x}_i,t)X(\mathbf{x}_j,t), \tag{50}$$

$$\Sigma^A_{YY}(i,j) = \frac{1}{K}\sum_{t=1}^{K} Y(\mathbf{y}_i,t)Y(\mathbf{y}_j,t), \tag{51}$$

$$\Sigma^A_{XY}(i,j) = \frac{1}{K}\sum_{t=1}^{K} X(\mathbf{x}_i,t)Y(\mathbf{y}_j,t). \tag{52}$$

$$\tag{53}$$

The cross-covariance matrix in the EOF space is

$$\tilde{\Sigma}_{XY} = [\langle X_n Y_m \rangle]_{p\times q}\,. \tag{54}$$

Its transpose is

$$\tilde{\Sigma}'_{XY} = \tilde{\Sigma}_{YX}. \tag{55}$$

The maximization of the correlation ρ leads to the following eigenvalue problem for $\tilde{\mathbf{u}}$:

$$\tilde{\Sigma}_{XY}\tilde{\Sigma}_{YX}\tilde{\mathbf{u}} = \rho^2\tilde{\mathbf{u}}. \tag{56}$$

Similarly, the eigenvalue problem for $\tilde{\mathbf{v}}$ can be derived as well:

$$\tilde{\Sigma}_{YX}\tilde{\Sigma}_{XY}\tilde{\mathbf{v}} = \rho^2\tilde{\mathbf{v}}. \tag{57}$$

In practical computing, the cross-covariance matrix in the EOF space is calculated by time average

$$\tilde{\Sigma}_{XY} \approx \frac{1}{K}\sum_{t=1}^{K} X_n(t)Y_m(t), \tag{58}$$

where $X_n(t)$ is given by one of the following three expressions

$$X_n(t) \;=\; \frac{1}{\sqrt{\lambda^X_n}}\left(\mathbf{e}^X_n\right)' \mathbf{X}_A$$

$$= \sum_{i=1}^{J_X} \frac{1}{\sqrt{\lambda_n^X}} \mathbf{e}_n^X(i) X(\mathbf{x}_i, t) \sqrt{A_i}$$

$$= \sum_{i=1}^{J_X} \frac{1}{\sqrt{\lambda_n^X}} \psi_n(\mathbf{x}_i) X(\mathbf{x}_i, t) A_i. \quad (59)$$

Using what expression depends on the available types of EOFs. The quantity $Y_m(t)$ can be computed in the similar way.

4 MULTIVARIATE REGRESSION FOR ANOMALY PREDICTION

Our prediction method is a linear regression based upon the data of the predictor and predictand. The predictand is a linear combination of the predictor. The predictand and predictor can be defined either on points or regions, although they are often defined on grid points. Of course, the regression is meaningful only when there is an approximate linear relationship between the predictor and predictand. The regression is statistically significant only when the scatter diagram of the data are not too far away from the regression line (or the regression plane in the case of multivariate). This condition is equivalent to that the correlation between the two fields is large.

In the case of single-variate, the regression line

$$y = bx + x_0$$

is statistically significant, only when the data point (x_i, y_i) are close to the line. The slope is $b = \rho\sigma_y/\sigma_x$, where ρ is the sample correlation, σ_x and σ_y are the sample standard deviations of the x-data and y-data, respectively. How close the points are to the regression line is measured by the sample correlation, ρ, computed from the data. The closer the absolute value of ρ is to one, the closer the points to a line, and consequently, the more significant the regression. For example, when having 15 data points, the 5% significance level requires the correlation not less than 0.514.

Our climate prediction uses multivariate regression, which is a linear approximation to a nonlinear relationship. The correlation between the two fields is defined in the previous sections, and computed by the CCA algorithms. Our linear approximation is optimal in the sense of the maximization of the correlation between the two fields.

Our linear prediction in the physical space can be written as

$$Y(\mathbf{y}_i, t') = \sum_j B_{ij} X(\mathbf{x}_j, t). \tag{60}$$

Here, the predictand is at time $t' = t + \Delta t$, the predictor is at time t, the "slope" matrix B is to be found, and the "intercept" vector is set to be zero since both Y and X are anomalies. We will use the canonical variables and canonical correlation patterns, together with EOFs, to express the fields and then find the "slope" matrix B. The correlation between the X-field and Y-field is the key parameter in the regression.

If the truncation order $p > q$, then the correlation between $U^{(k)}(t)$ and $V^{(k)}(t)$ is zero when $k > q$. The same is true for the case $q > p$. Thus, we always make EOF truncation order for X and Y be the same: $p = q$. The CCA eigenvalue problem has p non-zero eigenvalues. When $s = 2p - K + 1 > 0$ (where K is the length of the data stream), the first s eigenvalues are equal to one, which implies a perfect linear relationship that is certainly impossible. Thus, choosing a large p may result in over-fit. In this paper, we always choose $p < (K-1)/2$ and none of the eigenvalues is equal to one.

The p eigenvectors are mutually orthogonal. The canonical variables are

$$U^{(k)}(t) = \sum_{n=1}^{p} u_n^{(k)} X_n(t), \tag{61}$$

$$V^{(k)}(t') = \sum_{n=1}^{q} v_n^{(k)} Y_n(t'), \quad k = 1, 2, \cdots, p. \tag{62}$$

Now, the sign of $\tilde{\mathbf{u}}^{(k)}$ needs to be determined to guarantee the positivity of ρ_k. One can calculate ρ using

$$\gamma_k = \frac{1}{K} \sum_{t=1}^{K} U^{(k)}(t) V^{(k)}(t + \Delta t).$$

If γ_k is positive, then $\rho_k = \gamma_k$ and $\tilde{\mathbf{u}}^{(k)}$ does not change sign. Otherwise, $\gamma_k < 0$, then $\rho_k = -\gamma_k$ and $\tilde{\mathbf{u}}^{(k)}$ is changed to $-\tilde{\mathbf{u}}^{(k)}$. After this operation, the correlation at each mode (k) is non-negative.

It is worthwhile to note that the operations above are among the dimensionless variables. Because the matrices $[u_n^{(k)}]_{p\times p}$ and $[v_m^{(k)}]_{p\times p}$ are orthogonal matrices, their inverses are equal to their transposes. Such an inverse operation for a matrix is physically meaningful only when the matrix is dimensionless.

The inverse of the matrix equations (61) and (62) leads to

$$X_n(t) = \sum_{k=1}^{p} u_n^{(k)} U^{(k)}(t), \tag{63}$$

$$Y_n(t') = \sum_{k=1}^{p} v_n^{(k)} V^{(k)}(t'). \tag{64}$$

The field X in the spectral space is expressed by the sum of the products of the canonical variable $U^{(k)}$ and eigenvector $\tilde{\mathbf{u}}^{(k)}$. The eigenvector $\tilde{\mathbf{u}}^{(k)}$ is thus called the canonical pattern in spectral space. Its physical space correspondence is $\mathbf{u}_A^{(k)}$ given by formula (48).

Please note that, for the purpose of one-tier prediction, the predictand time t' is larger than the predictor time t by Δt and, thus, the cross-covariance matrix $\tilde{\Sigma}_{XY}$ in this section is a lagged cross-covariance. The lag time is $t' - t = \Delta t$.

The vectors $(U^{(1)}, U^{(2)}, \cdots, U^{(p)})'$ and $(V^{(1)}, V^{(2)}, \cdots, V^{(p)})'$ are all in p-dimensional vector space. There must exist a unique transform matrix which transform one to the other. Let this transformation be written as

$$V^{(k)} = \sum_{l=1}^{p} F_{kl} U^{(l)} + b_k, \quad k = 1, 2, \cdots, p. \tag{65}$$

The formulas

$$\langle V^{(k)} U^{(l)} \rangle = \rho_k \delta_{kl}, \tag{66}$$

$$\langle U^{(k)} \rangle = 0, \quad \langle \left(U^{(k)} \right)^2 \rangle = 1, \tag{67}$$

$$\langle V^{(k)} \rangle = 0, \quad \langle \left(V^{(k)} \right)^2 \rangle = 1 \tag{68}$$

imply that

$$V^{(k)}(t') = \rho_k U^{(k)}(t), \quad k = 1, 2, \cdots, p. \tag{69}$$

The prediction is therefore made for $t' = t + \Delta t$ when the predictor at t is known:

$$X(\mathbf{x}, t) -> X_n(t) -> U^{(k)}(t) -> V^{(k)}(t') -> Y_n(t') -> Y(\mathbf{y}, t').$$

The equation (69) is thus the main prediction equation. The predictand at t' is

$$\hat{Y}(\mathbf{y}, t') = \sum_{n=1}^{p} \left(\sum_{k=1}^{p} v_n^{(k)} \rho_k U^{(k)}(t) \right) \sqrt{\lambda_n^Y} \phi_n(\mathbf{y}). \tag{70}$$

This is the one-tier prediction scheme: knowing the predictor at time t and predicting the predictand at $t' = t + \Delta t$. The CCA scheme can also be applied to the so called two-tier forecasting: forecasting the predictor from time t to time $t' = t + \Delta t$ by a method and predicting the predictand at time t' from the predictor also at time t' by another method. CCA can also be used as the latter method. The CCA scheme is exactly the same, except that the lagged cross covariance matrix should be replaced by cross-covariance matrix without time lag, i.e. $\Delta t = 0$.

5 ESTIMATE OF THE PREDICTION ERROR

This section estimates the mean square error (MSE) of prediction. The importance of this error estimate is three fold. First, it gives an instantaneous assessment of the forecasting quality. The smaller the error, the more accurate the forecast. Second, the MSE is directly related to the correlation skill of the prediction which is often used to assess the quality of a prediction. Third, the MSE errors are needed to compute the optimal weights for combining the multiple forecasts to form an optimal ensemble forecast.

The CCA scheme requires the fields to be stationary. The computing algorithm requires the fields to be ergodic, i.e. the ensemble mean can be replaced by time mean for a long time. Our estimated error, as an ensemble value, cannot reflect the fluctuation of the fields according to time. However, if the data used in the prediction are changed, then the error should reflect this change. One may regard an observational network as a filter. When a filter is changed and the filtered results should change accordingly. Thus, the time dependence of the error is on the observational data and the data processing method, not on the intrinsic fluctuations of the fields. Thus, for a given observational network and a given data-processing method, our forecasting error is time independent.

The error field is

$$\epsilon^2(\mathbf{y}) = \langle \Big(Y(\mathbf{y}, t') - \hat{Y}(\mathbf{y}, t') \Big)^2 \rangle, \tag{71}$$

where Y is the unknown true field to be predicted, and $\hat{Y}$ is the prediction that approximates Y. The prediction skill of $\hat{Y}$ is often measured by its correlation with the original field Y, defined by

$$r = \frac{\langle Y\hat{Y} \rangle}{\sqrt{\langle Y^2 \rangle \langle \hat{Y}^2 \rangle}}. \tag{72}$$

From above two equations one can derive

$$r = \frac{1}{2} \left[\sqrt{\frac{\langle Y^2 \rangle}{\langle \hat{Y}^2 \rangle}} + \sqrt{\frac{\langle \hat{Y}^2 \rangle}{\langle Y^2 \rangle}} - \frac{\epsilon^2}{\sqrt{\langle Y^2 \rangle \langle \hat{Y}^2 \rangle}} \right]. \tag{73}$$

Several special cases are interesting. Case 1 is the perfect prediction: $\epsilon = 0$ implies $r = 1$. Case 2 is the worst wrong prediction: $\epsilon^2 = \langle Y^2 \rangle + \langle \hat{Y}^2 \rangle$ implies $r = 0$. And case 3 is the random prediction: $\epsilon^2 = \langle Y^2 \rangle$ implies $r = (1/2)\sqrt{\langle \hat{Y}^2 \rangle / \langle Y^2 \rangle}$. A sub-case of this is $\hat{Y}^2 = 0$, which is a useless forecast. In other sub-cases, it is unlikely that $\langle \hat{Y}^2 \rangle$ is very small, hence the correlation r is also very small.

The MSE can be estimated by using EOFs. The error can be written into two parts:

$$\epsilon^2(\mathbf{y}) = \epsilon_q^2(\mathbf{y}) + \epsilon_R^2(\mathbf{y}), \tag{74}$$

where

$$\epsilon_q^2(\mathbf{y}) = \sum_{m=1}^{q} \epsilon_m^2 \phi_m^2(\mathbf{y}) \tag{75}$$

is MSE computed from the first q modes with

$$\epsilon_m^2 = \langle \Big(Y_m(t') - \hat{Y}_m(t') \Big)^2 \rangle, \tag{76}$$

and the residual $\epsilon_R^2(\mathbf{y})$ is the contribution of higher modes to the MSE:

$$\epsilon_R^2(\mathbf{y}) = \sum_{m=q+1}^{\infty} \lambda_m \phi_m^2(\mathbf{y}). \tag{77}$$

Apparently, this residue error depends on the truncation order, which, in turn, depends on the resolution of data and the spatial scale of the field. The higher the data resolution, the larger the q value. In practice, one uses discrete data and obtains only a finite number of modes, say, q'. The prediction uses the first q modes among the $q' > q$ modes. Then, the residual error may formally written as

$$\sum_{m=q+1}^{q'} \lambda_m^Y \phi_m^2(\mathbf{y}).$$

This q' is usually chosen to be K, the length of the data stream. However, when the data streams are not sufficiently long, they cannot resolve the EOFs of very high order. Because the higher order eigenvalues are close to each other, the computed EOFs can be very much distorted from the true ones (See North's rule-of-thumb, North et al., 1982). Thus, the above estimate of residual error can be unreliable if the data streams are not sufficiently long.

Usually, the q modes of the predictand should resolve at least 50% total variance. The residual error should be smaller than or comparable to the prediction error of the first q modes. With the omission of the residual error in the computation, the error $\epsilon_q^2(\mathbf{y}, t')$ computed based on the first q modes can be regarded as the lower bound of the MSE and its double may be regarded as the upper bound of the MSE.

The average MSE of the error field is the integral of $\epsilon^2(\mathbf{y}, t')$

$$\epsilon_A^2 = \frac{1}{A} \int_{\Omega_Y} d\Omega_Y \, \epsilon^2(\mathbf{y}, t'), \tag{78}$$

where A is the total area of the domain Ω_Y. When using the first q modes, the average MSE error is

$$\epsilon_A^2 = \frac{1}{A}\sum_{m=1}^{q}\epsilon_m^2. \tag{79}$$

This MSE's unit is still [Y unit]^2, as one expected.

The error for the mth mode is computed as following

$$\begin{aligned}
\epsilon_m^2 &= \lambda_m^Y\langle\left(Y_m(t') - \hat{Y}_m(t')\right)^2\rangle \\
&= \lambda_m^Y\langle\left(\sum_{k=1}^{q} v_m^{(k)}V^{(k)}(t') - \sum_{k=1}^{q} v_m^{(k)}\rho_k U^{(k)}(t')\right)^2\rangle \\
&= \lambda_m^Y\langle\left[\sum_{k=1}^{q}\left(V^{(k)} - \rho_k U^{(k)}\right)v_m^{(k)}\right]^2\rangle \\
&= \lambda_m^Y\sum_{k=1}^{q}(1-\rho_k^2)\left(v_m^{(k)}\right)^2.
\end{aligned} \tag{80}$$

The error is scaled down for higher EOF modes of the Y-field by its eigenvalues λ_m^Y since $\lambda_m^Y \to 0$ as $m \to \infty$, but scaled up by the CCA eigenvalues ρ_k^2 because $1-\rho_k^2 \to 1$ as $k \to \infty$. If the correlations ρ_k^2 are large, say, close to one, then $\epsilon_m \approx 0$. If the predictability is low, then the correlation is small. If the largest correlation is even less than 0.4, the factor $1-\rho_1^2$ is larger than 0.84, which is pretty large. If both the predictor and predictand are spatial white noise, then there is no correlation between the two fields, and $\rho_k^2 = 0$ and

$$\epsilon_m^2 = \lambda_m^Y,$$

and

$$\epsilon_A^2 = \sum_m \lambda_m^Y = \langle \int_{\Omega_Y} d\Omega_Y\ Y^2(\mathbf{y},t)\ \rangle.$$

Namely, the prediction error is equal to the total variance of the field and the prediction skill is zero.

6 OPTIMAL ENSEMBLE OF MANY FORECASTS

Different data sets of predictor and predictand may yield many forecasts, such as the forecasts of precipitation using SST, soil moisture, SLP, SLP, or SST from different ocean basins, etc. An optimal weight needs to be assigned to each forecasting result at every grid point in order to form an optimal ensemble forecast. This is similar to the superensemble idea proposed by Krishnamurti et al. (2000).

This section will show the method to find the optimal weights and the resulting prediction error. The spectral method is again used in our computing. The use of the spectral method requires stationarity assumption. Because of the requirement, which may be true in a short period, the forecasting naturally deteriorates rapidly as the lead time of the prediction increases.

The spectral representation of the predictand is

$$Y(\mathbf{y},t)=\sum_{n=1}^{\infty}\sqrt{\lambda_n^Y}Y_n(t)\phi_n(\mathbf{y}). \tag{81}$$

Similarly, the forecaster's representation is

$$\hat{Y}(\mathbf{y},t)=\sum_{n=1}^{\infty}\sqrt{\lambda_n^Y}\hat{Y}_n(t)\phi_n(\mathbf{y}). \tag{82}$$

Suppose that there are H forecasts. The ensemble forecast for the principal component $Y_n(t)$ is

$$\hat{Y}_n(t)=\sum_{h=1}^{H}w_n^{(h)}\hat{Y}_n^{(h)}(t), \tag{83}$$

where $w_n^{(h)}$, to be determined, is the weight for model k and mode n. The weights satisfy a constraint

$$\sum_{h=1}^{K}w_n^{(h)}=1, \tag{84}$$

but they do not have to be all positive.

The MSE forecasting error is a function of $\mathbf{y}$ and is written as

$$\begin{aligned}E^2(\mathbf{y})= &\quad\langle\Big(Y(\mathbf{y},t)-\hat{Y}(\mathbf{y},t)\Big)^2\rangle\\ &=\langle\left[\sum_{n=1}^{\infty}\left(Y_n(t)-\sum_{h=1}^{H}w_n^{(h)}\hat{Y}_n^{(h)}(t)\right)\sqrt{\lambda_n^Y}\phi_n(\mathbf{y})\right]^2\rangle.\end{aligned} \tag{85}$$

The principal components of different modes are independent. The errors $Y_n-\hat{Y}_n$ $(n=1,2,3,\cdots)$ are assumed to be independent. Then, the error function becomes

$$E^2(\mathbf{y})=\sum_n E_n^2\phi_n^2(\mathbf{y}), \tag{86}$$

where

$$E_n^2=\lambda_n^Y\langle\left(Y_n(t)-\sum_{h=1}^{H}w_n^{(h)}\hat{Y}_n^{(h)}(t)\right)^2\rangle \tag{87}$$

is the MSE error for forecasting the nth principal component $Y_n(t)$.

The forecasting errors are caused by noisy data or inaccuracy of the forecasting model. For each EOF coefficient, the forecasting errors for different prediction models are assumed uncorrelated. With this assumption, one then has the following

$$\begin{aligned} E_n^2 &= \lambda_n^Y \langle \left(\sum_{h=1}^{H} w_n^{(h)} \left(Y_n(t) - \hat{Y}_n^{(h)}(t) \right) \right)^2 \rangle \\ &= \lambda_n^Y \sum_{h=1}^{H} \left(w_n^{(h)} \right)^2 \left(\epsilon_n^{(h)} \right)^2, \end{aligned} \tag{88}$$

where

$$\left(\epsilon_n^{(h)} \right)^2 = \langle \left(Y_n(t) - \hat{Y}_n^{(k)}(t) \right)^2 \rangle. \tag{89}$$

Minimization of ϵ_n^2 expressed as above under the constraint (84) leads to the optimal weights. The weight for a model is inversely proportional to the model's forecasting error. Namely,

$$w_n^{(h)} = \frac{1/(\epsilon_n^{(h)})^2}{\sum_{l=1}^{H} 1/(\epsilon_n^{(l)})^2}. \tag{90}$$

Then the resulted MSE error for the mode n is

$$E_n^2 = \left(\sum_{h=1}^{H} 1/(\epsilon_n^{(h)})^2 \right)^{-1}. \tag{91}$$

That the errors from different prediction models are independent is an important assumption. The different models usually mean different predictors, but the same prediction scheme. Thus, to ensure the independence of the errors, one should choose the predictors which are not strongly correlated. This is intuitively correct, since the information from the similar sources is redundant.

A special case is when all the models have the same amount of errors, $(\epsilon_n^{(0)})^2$. In this case, the optimal weights are all equal to $1/H$ and the resulting error is

$$\epsilon_n^2 = \frac{(\epsilon_n^{(0)})^2}{H}. \tag{92}$$

This is the conventional error-formula of homogeneous statistics.

The method of finding optimal weights here is slightly different from that of Krishnamurti et al. (2000), though the purposes of the two methods are the same: optimal combination of multiple forecasts. Krishnamurti et al. (2000) used the multivariate regression to determine the weights, and they did not require the sum of the weight to be one. Since their regression is for the best fit of the linear superensemble model to data, the forecasted expected values

are approximately the same as the observed expected values and hence the constraint of the sum of all the weights equal to one is not needed. But in our case, the constraint is needed to guarantee that the forecasted expected values is approximately the same as the observed expected values. Otherwise, it may happen that the weights are too large, and, consequently, the forecasted expected value is out of the range of possible climate and hence the forecast is not valid. Despite the difference between the present method and that of Krishnamurti et al. (2000), the results are actually close. We applied the formula (91) to the Y-component of their Fig. 3(b), the formula (91) provides an RMSE 1.7, which is about the same as that of Krishnamurti et al. (2000).

7 CHARACTERISTICS OF THE US PRECIPITATION

To predict the seasonal precipitation over the US, it is helpful to describe its basic characteristics. Higgins et al. (1996) made an atlas of the US monthly precipitation, which is a valuable reference for forecasting research. The US seasonal precipitation is influenced by many factors, including sea surface temperature (SST), sea level pressure (SLP), soil moisture, wind direction, wind speed, and relative humidity. The cold winter-air mass from Canada often trigger heavy snow fall along the US snow-belt. The warmer SST over the eastern Pacific often results in a positive precipitation anomaly along the coast of California, particularly during an El Nino event. The anomaly of the Atlantic SST is often the energy source of the Atlantic hurricane, which, when it lands in the east coast of the US, makes a significant contribution to a monthly precipitation of the eastern states, including North and South Carolinas, Virginia, and sometimes, Texas, Georgia, New York, and Maine.

The US annual precipitation has two centers. A center of a larger area is around the central Gulf Coast states, including Louisiana, Mississippi, and Alabama. This area's precipitation almost uniformly distributed from January to December, with slightly lower amount in October and slightly higher amount in July and August. This area's precipitation is mainly influenced by the low pressure system over the northern part of the Gulf of Mexico. The winter precipitation is affected by winter cyclones. The spring precipitation is shifted a little toward the west, over Arkansas and Oklahoma, where the Canadian cold air often brings down the moist transported from the Gulf of Mexico and triggers severe convective storms. The PRE-STORM experiment of the summer of 1985 in Oklahoma was designed

to measure this type of precipitation. The summer precipitation is also influenced by the weather systems of the Atlantic, such as large scale hurricanes.

Another center of a smaller area is the northwest US, including only the Pacific Northwest coast states of Washington and Oregon. This region is small, but has the largest amount of precipitation comparing to other regions of the US. This area's precipitation, different from that of the southeast center, is highly non-uniform from January to December. It has a strong seasonal cycle, with maximum precipitation in winter, minimum in summer, and moderate in spring and fall. The winter precipitation is mainly caused by the confront of the northwest cold air jet from Alaska and the westerly Pacific jet. The latter carries much of water vapor. Many winter days have more than 5 mm/day precipitation rate and make the area the wettest region of the conterminous US. The westerly Pacific jet becomes weaker in summer and shifts to higher latitude, and hence it has stronger influence to British Columbia than Washington. This shift dramatically reduces the supply of water vapor to the Washington and Oregon area and hence reduces the precipitation amount.

Of course, interannual variations of SST also influence the US precipitation (Ropelewski and Halpert, 1986; and others). During the El Nino event, the precipitation of California has a strong positive winter anomaly, and that of the southeast US has a strong negative spring anomaly. As the El Nino signal is the strongest in December, the CCA should yield a high skill over California if the December's SST is used to predict the January precipitation. The CCA should also yield a high skill over the southeast US when the November-December-January SST is used to predict the February-March-April precipitation. The cold Pacific SST anomaly, such as La Nina, results in a negative winter precipitation anomaly in the southwest US, while the cold events' influence in other regions and seasons is pretty much scattered.

With the above information, one can be certain that the US precipitation is correlated to the surface temperature (particularly, the SST) and pressure field. These properties provide the predictability and our optimal forecasting scheme is designed to reflect the predictability and to yield high quality prediction. Of course, our method can only reflect an linear approximation of the nonlinear relation between the SST and US precipitation. Although the linear approximation has limitations, it has already provided meaningful results. Barnston and Smith (1996) used the BP-CCA to predict the global precipitation using the global SST and found significant prediction skill over North America at least in the winter seasons.

8 DATA SETS AND THE RESULTS OF THE US PRECIPITATION FORECASTING

8.1 Data sets

Our predictor's data set is the historical monthly SST reconstructed by Smith et al. (1996). The spatial coverage is from 45°S to 69°N. The spatial resolution is 2° × 2°. The temporal coverage is from January 1951 to December 2000. The temporal resolution is a month. The data are available via ftp at: ftp://ftp.ncdc.noaa.gov/pub/data/ocean/rsst/. Our linear regression scheme cannot resolve the nonlinear process imbedded in small scales. Thus, the 2°×2° SST data are further averaged into a data set of 4-degree latitude by 6-degree longitude. This average removes the noise of smaller scales.

The predictand's data set is based upon the latest version of the Xie-Arkin global land monthly precipitation data set as reported in Chen et al. (2001). The temporal coverage of the precipitation data is from January 1948 to December 2000, but only the part overlapped with the reconstructed SST period is used in our forecasting calculation, namely, January 1951 - December 2000. The spatial resolution is 2.5° × 2.5°. The gridded data are resulted from the interpolation of over 17,000 stations collected in the Global Historical Climatology Network (GHCN) Version 2 and the Climate Anomaly Monitoring System (CAMS). The data are available via ftp at: ftp.ncep.noaa.gov/pub/precip/50-yr.

The anomaly data are centered around to the 1961-1990 climatology. These anomaly data are further detrended and standardized before they go into the EOF computing. After the forecasting on the standardized data, the forecasting results (for dimensionless quantities) are converted back to the anomaly data with dimension. However, the conversion uses only the mean and standard deviation, and the trend is not included. The forecasting results are then compared with the observed detrended anomaly field to assess the forecasting accuracy.

8.2 Forecasting results

The 50 years (1951-2000) SST and precipitation data are used in our forecasting experiments. Hence 49 years data are used for computing EOFs for both SST and precipitation, and these data are used to predict the remaining year's precipitation. Our scheme requires that the lengths of the data streams for both SST and precipitation be the same. The reconstructed

SST data are only for the preiod of 1951-2000, the precipitation data before 1951, although available, are not used in our forecasting experiments. EOFs of the US precipitation are presented for two seasons: DJF (winter, Fig. 1) and JJA (summer, Fig. 2). For both precipitation and SST fields we show only the first six EOF modes. The number of modes used for actual forecasting is ten or another number.

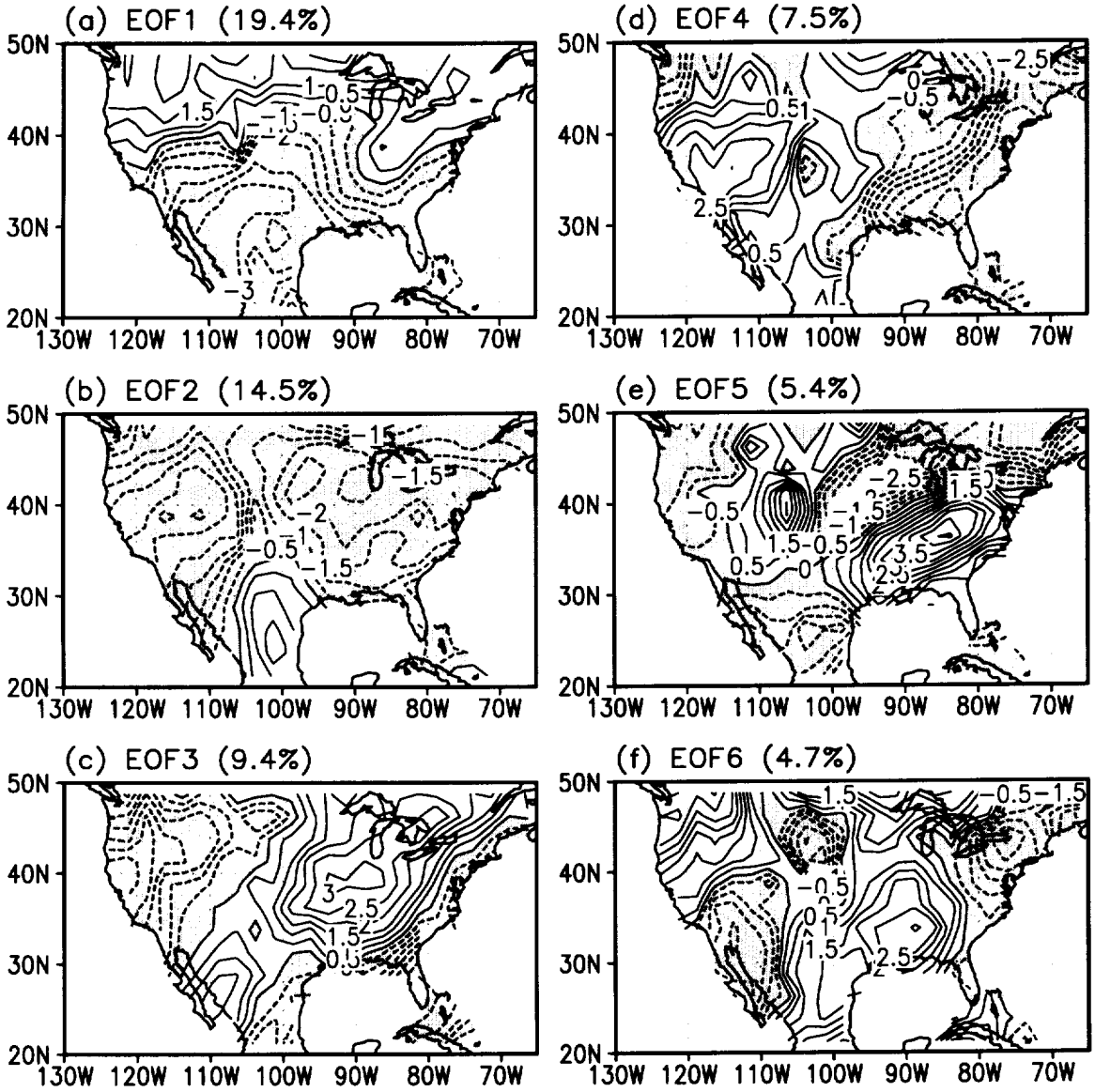

Figure 1: First six EOFs of US precipitation in DJF. The computation used the three-month mean precipitation data, which had been detrended and standardized.

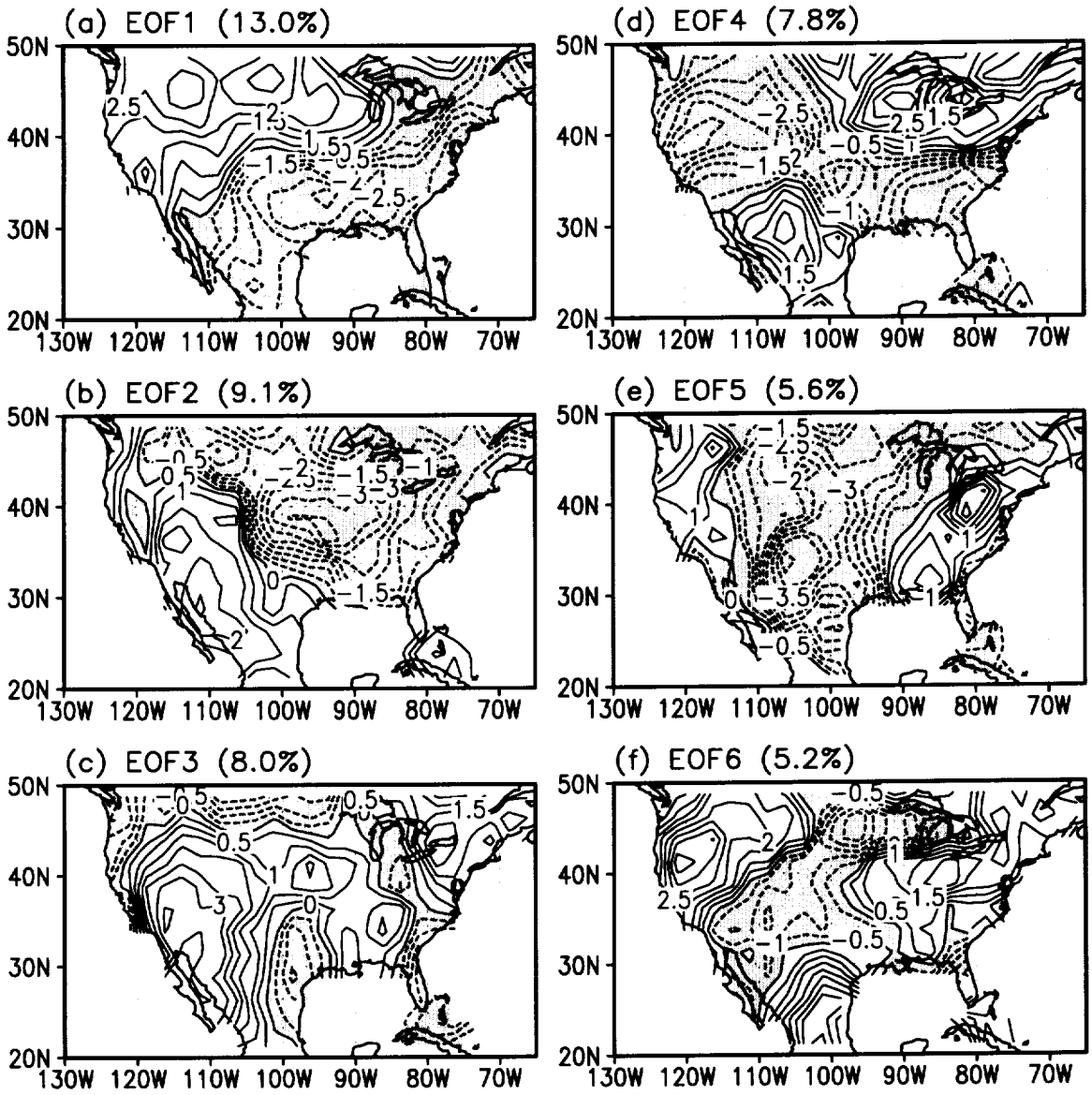

Figure 2: Same as Fig. 1, except for the summer season JJA.

It is well known that the characteristics of summer precipitation is very much different from that of winter. The EOFs clearly reflect the difference. For example, the DJF EOF1 demonstrates a strong north-south pattern, while the JJA EOF1 supports a northwest-southeast pattern. The differences are also reflected in higher modes.

The CCA patterns are linear combinations of the EOFs. The interesting part is that the first two winter CCA patterns can sometimes be highly mode-selective, preferring only a few from the six retained EOF modes. When performing the SON (SST field)- DJF (precipitation) CCA analysis, the first SST CCA pattern is mainly contributed by 0.75 EOF1

and 0.61 EOF2. Since the first SST EOF is the El Nino mode, the first SST CCA pattern looks like the El Nino mode plus some North Pacific oscillation (see Fig. 3). The contribution from other EOFs is negligible. The first precipitation CCA pattern is mainly contributed by 0.68 EOF1, 0.38 EOF2, and 0.34 EOF5 (Fig. 3). The contribution from other modes is small. The northwest positive is mainly contributed by EOF1.

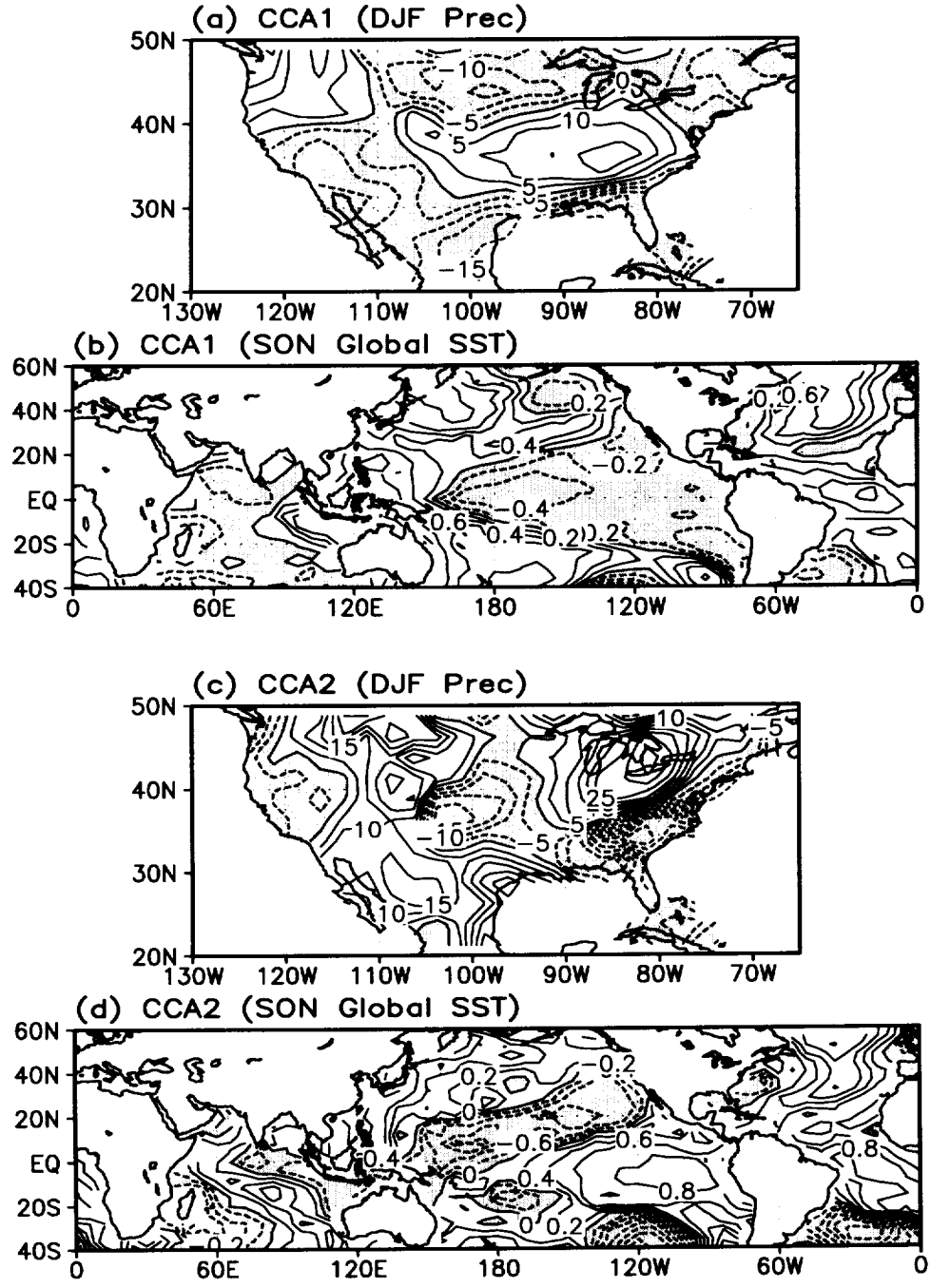

Figure 3: First two canonical correlation patterns for SON SST and DJF US precipitation.

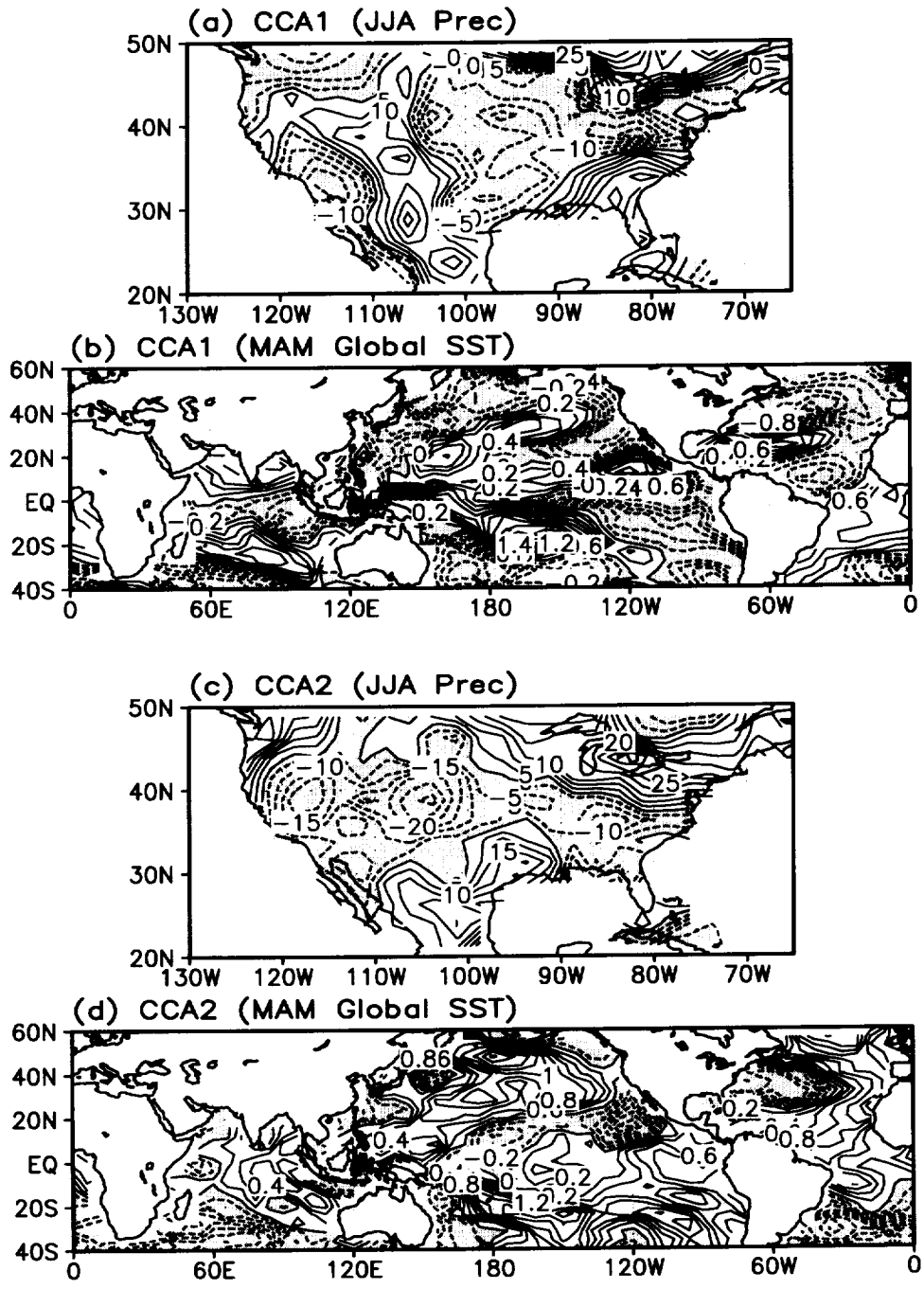

Figure 4: First two canonical correlation patterns for MAM SST and JJA US precipitation.

With this pair of CCA patterns, the CCA correlation between the SST and precipitation field (i.e., ρ_1) is 0.87, a pretty large value indicating that a good forecasting result is expected. However, the summer CCA patterns are more uniformly distributed among the ten retained EOFs and do not have apparent dominant modes. For example, the first precipitation CCA

pattern selects not only EOF1 and EOF2 , but also EOF6, EOF7, EOF8 and EOF9 (Fig. 4). Comparing Fig. 4(a) and Fig. 3(a), one can see that DJF's first canonical pattern is much more coherent than that of JJA. The same is true for the SST canonical patterns (see Figs. 3(b) and 4(b)). This reflects the difficulty of predicting the summer precipitation.

The 2000 DJF precipitation is predicted by the 1999 SON SST. The observed precipitation anomaly is shown in Fig. 5(a) and the predicted anomaly in Fig. 5(b). The predicted and the observed anomalies agree in most of the southeast area. The spatial pattern correlation, defined by the formula (93) below, between the predicted and observed is 0.47. The corresponding pattern correlation is higher for El Nino or La Nina years, but the year 2000 was neither an El Nino nor an La Nina year.

The DJF precipitation forecasts are made for every year from 1951 to 2000 by using the previous season's (i.e. SON) SST. The pattern correlation and the Heidke skill score are shown in Fig. 6. Heidke score accounts for only the correct anomaly sign. Pattern correlation accounts for both sign and amplitude. Only the perfect forecast, with both correct sign and amplitude, has pattern correlation equal to one. Fig. 6 indicates that our forecasting results are reasonable.

The spatial pattern correlation is the correlation between the forecasted ($\hat{R}_i$) and observed anomalies (R_i) for a given season,

$$\gamma = \frac{\sum_i R_i \hat{R}_i}{\sqrt{\sum_i R_i^2}\sqrt{\sum_i \hat{R}_i^2}}. \tag{93}$$

When the pattern correlation reaches 0.4, the forecasted field has a good similarity to the observed. Among our 50 years of forecasts, 29 years have the pattern correlation equal to or greater than 0.4. Our maximum pattern correlation is 0.78, which indicates an extremely good forecast (see Fig. 6(a)).

The Heidke score (HS) is another commonly used index to indicate the prediction skill (van den Dool et al., 1997). The two-class HS is defined as

$$HS = \frac{H - E}{T - E} \times 100, \tag{94}$$

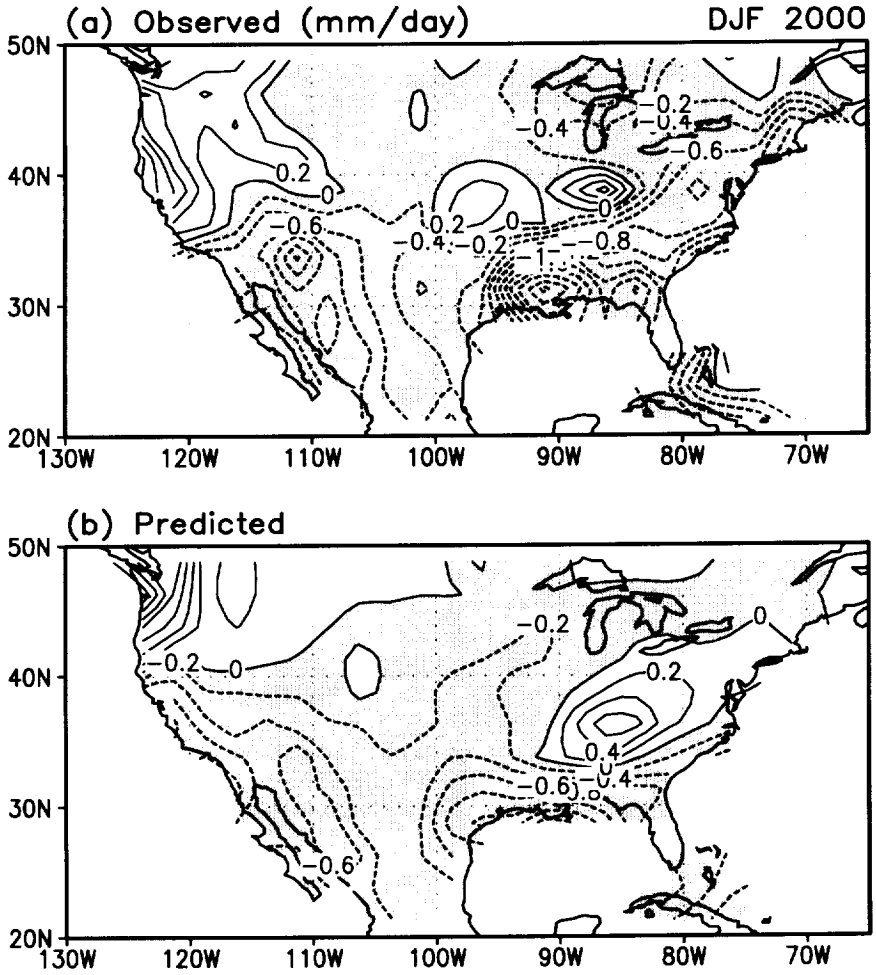

Figure 5: Observed and predicted precipitation field of DJF 2000. The prediction was made from the previous season SON SST. The time lag is hence one season. The unit is [mm]/[day].

where T is the total number of forecasts (i.e., the total number of grid points in our case), E is the expected value of correct random forecasts (i.e., $E = T/2$), and H is the actual number of correct forecasts. Since the two-class Heidke score is used, a forecast is defined as correct if the forecasted anomaly has the same sign as the observed anomaly. The maximal value of HS is equal to 100, when every forecast is correct. The HS score for random guessing is zero because $H = E$, hence the random forecasting model does not have a skill. The HS reaches its minimum value if every forecast has missed and the score is $-100 \times E/(T - E)$, which is equal to -100. Our forecasts show positive skill scores in 45 years in the total of 50 years. The maximum Heidke score is 76 (see Fig. 6(b)).

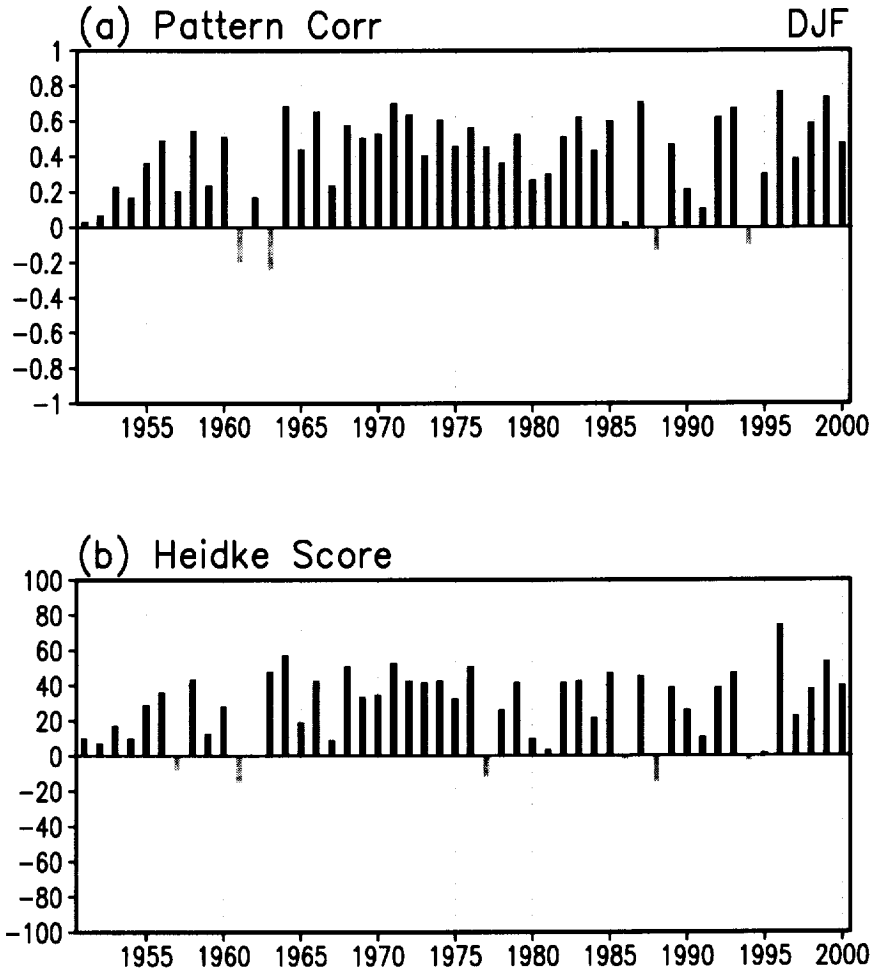

Figure 6: Forecasting skills of using previous season SON SST to predict DJF precipitation: (a) pattern correlation (defined by eq. (93)), and (b) Heidke score (defined by eq. (94)). The EOFs were computed using the data from 1951-1999.

The summer precipitation skill is consistently lower than that for winter. Among the 50 years of experiments, there are only 18 years whose the pattern correlations are equal to or greater than 0.4. The skill varies strongly with near-zero skill in many years, but the maximum skill can still be high. In our example, the maximum skill is 0.83. The pattern correlation for the 2000 prediction is very low and the forecast is not good. We have chosen to show a forecast of a bit less-than-average skill. This is 1997. The training period of the predictor and predictand is the 1951-1996. The 1997's JJA precipitation is predicted using the same year's MAM SST. The observed JJA precipitation is shown in Fig. 7(a). The predicted result is shown in Fig. 7(b). Comparing these two figures, one can see that the

forecasting skills are mainly from the Northeast and Southwest. In almost 50% of the US, the forecast has a wrong sign. Both the pattern correlation (0.18) between the forecasted and observed and the Heidke skill score are low (almost zero) (the most right bar in Fig. 8). This is a more detailed indication of the difficulty to forecast the summer precipitation.

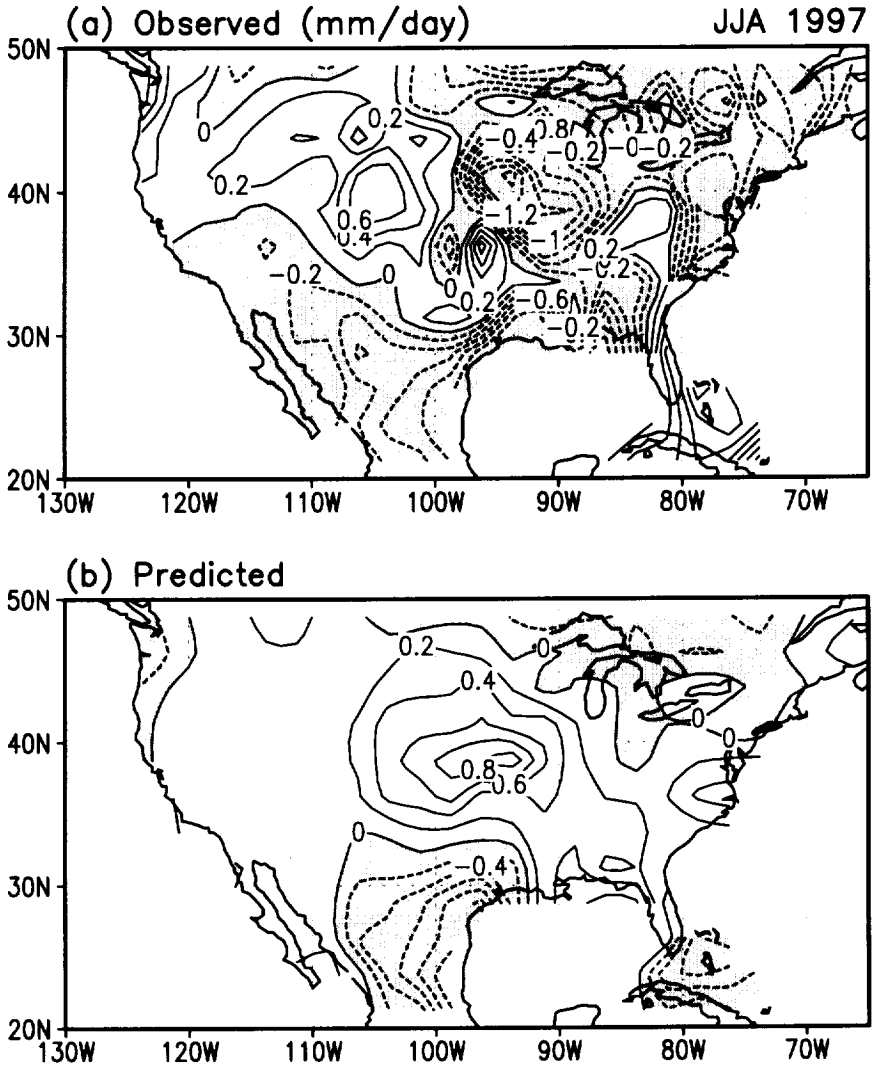

Figure 7: The observed and the forecasted 1997 JJA precipitation. The predictor was the 1997 MAM SST.

Of course, for actual forecast one does not know the observed precipitation. The reliability can be assessed by the expected value of prediction errors as discussed in Section 5. Fig. 9 shows the prediction root mean square errors (RMSE) for two seasons: DJF and JJA. The predictor is the previous season's SST. The errors measure the mean square difference between

the actual and forecasted precipitation. In the region where the precipitation variance is large, the error is also larger. The variance is proportional to the precipitation climatology. The DJF error map has a Northwest-Southeast pattern (Fig. 9(a)) and the JJA error error map has a Northeast-Southwest pattern. These patterns are consistent with US precipiation climatology (Higgins et al., 1996).

The RMSE error is the expected value. The regions of large errors are likely to have the wrong prediction. But for an individual forecast, the region of a large error may still have a good forecast.

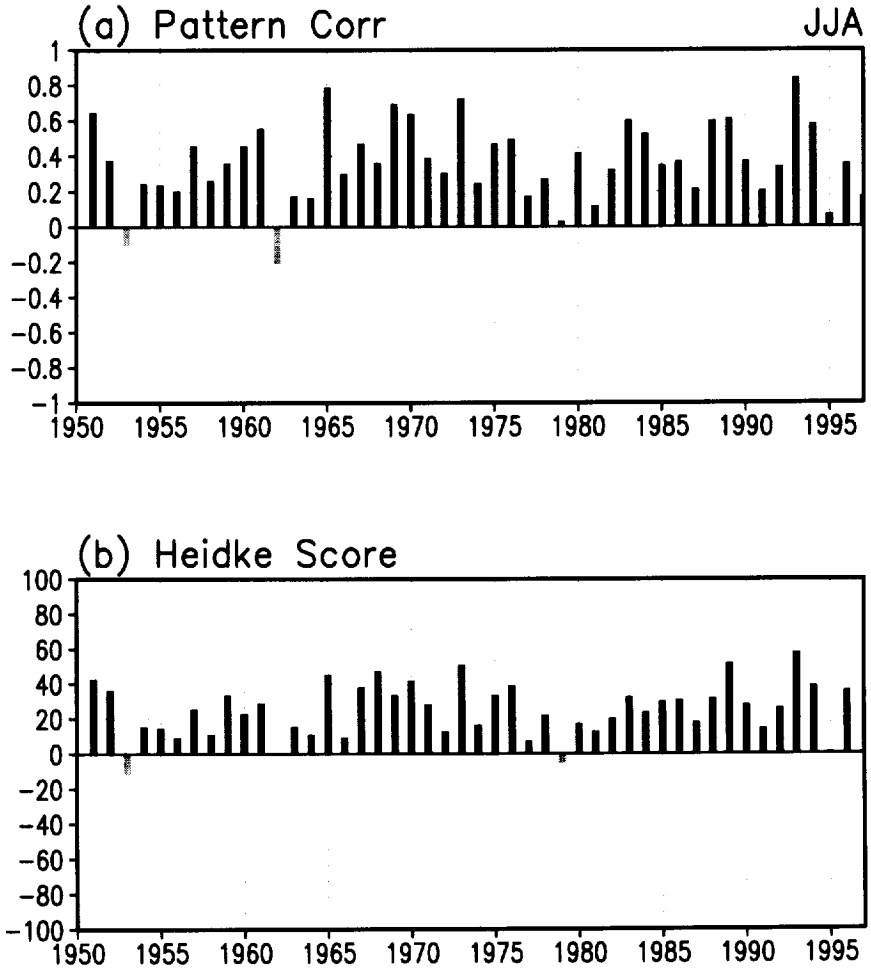

Figure 8: Forecasting skills of using MAM SST to predict JJA precipitation: (a) pattern correlation (defined by eq. (93)), and (b) Heidke score (defined by eq. (94)). The EOFs were computed using the data from 1951-1996.

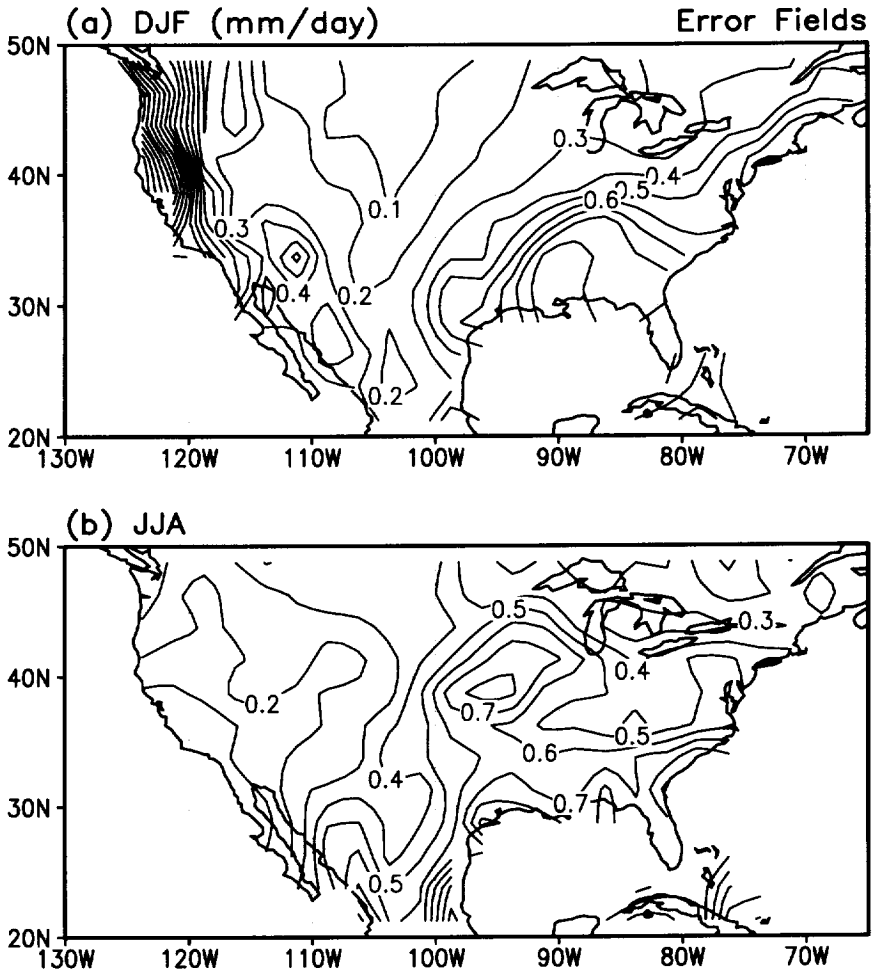

Figure 9: Expected values of the forecasting error when predicting the (a) DJF precipitation using the previous season SON SST, and (b) JJA precipitation using the same year's MAM SST.

Optimal combination of two forecasts may enhance the forecasting skill, but one has to be careful when doing so. Since both weights are positive, the optimal combination is an interpolation process. This process prefers two good forecasts. Thus, one should combine the two forecasts of reasonably high Heike scores. Fig. 10(c) shows the optimal combination of the 2000 DJF precipitation forecasts from the previous season's SST using the Pacific (Fig. 10(a)) and Atlantic oceans (Fig. 10(b)), respectively. The pattern correlation and Heidke score for Pacific forecast are 0.43 and 35, respectively. And those for Atlantic are 0.04 and 31. The optimally combined forecast is shown in Fig. 10(c). Its pattern correlation and Heidke scores become 0.47 and 40.

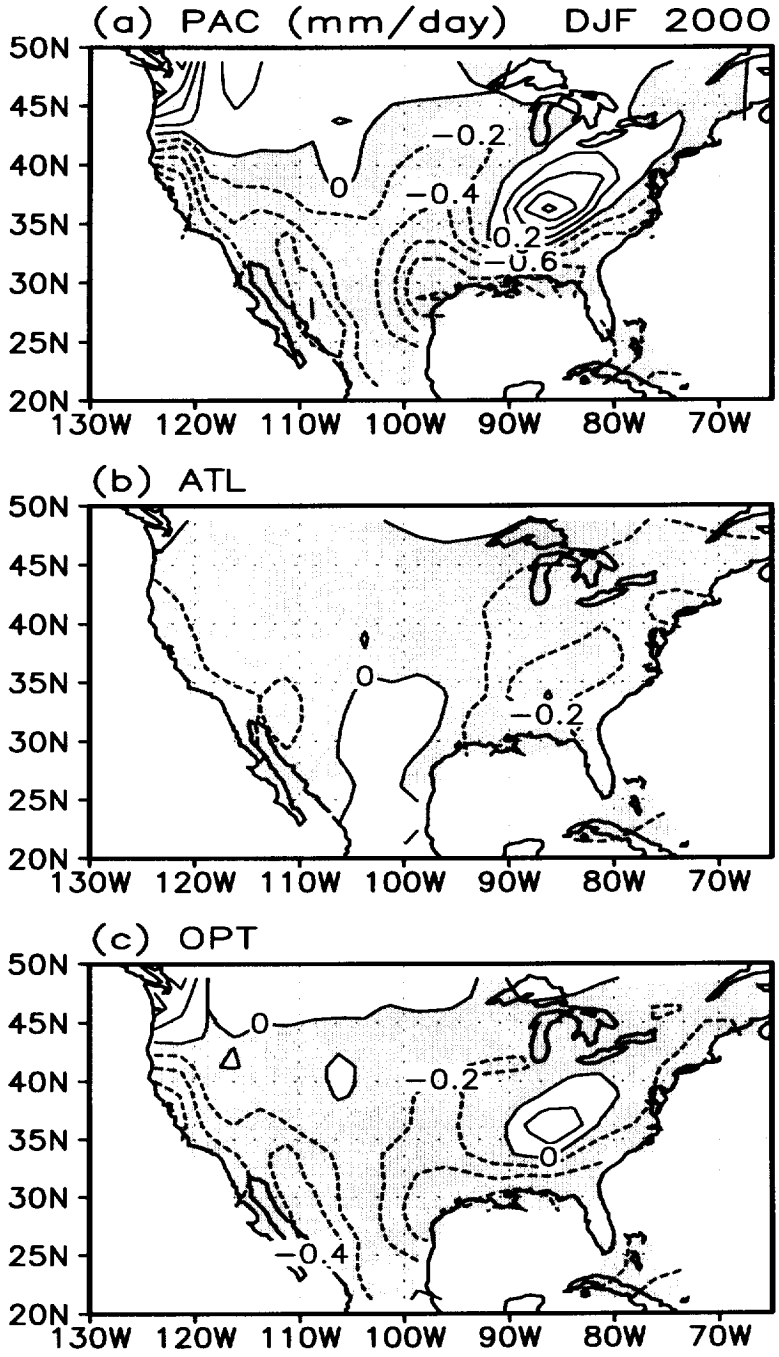

Figure 10: a) Predicted 2000 DJF precipitation using the previous season SON Pacific SST, (b) Predicted 2000 DJF precipitation using the previous season SON Atlantic SST, and (c)the optimal ensemble forecast, which is the optimal combination of the CCA forecasting results from (a) and (b). The spatial pattern correlations between the predicted and the observed (Fig. 5a) are 0.43 for (a), 0.04 for (b) and 0.44 for (c).

Compared with the observed precipitation in Fig. 5(a), the combined forecast has a better skill and is more similar to the observed than the two individual forecasts. If both forecasts are very good, the optimally combined forecast can have a much higher skill.

9 SUMMARY AND CONCLUSIONS

This paper contains a new development of the CCA forecasting scheme. This further development includes three parts. One is the inclusion of the area factor, and this procedure is very important to reduce the noise from the data of higher latitudes. The second is the error calculation formula, which estimates the mean square error of the forecasts and hence gives a quantitative quality assessment of the forecasting, without using cross validation. The error's spatial distribution reflects the physical interaction mechanisms between the predictor and predictand. The third is the optimal combination of the multiple forecasts. The third is possible only when the error estimates are made in the second.

Our forecast experiments show that the new ensemble canonical correlation scheme renders a reasonable forecasting skill. For example, when using September-October-November SST to predict the December-January-February precipitation of the next year, the spatial pattern correlation between the observed and predicted are positive in 46 years among the 50 years of experiments. The positive correlations are equal to or greater than 0.4 in 29 years, which indicates excellent performance of the forecasting model. The forecasting skill can be further enhanced when careful selection of ocean basins is made and more predictors are used.

With the results presented in this paper, we may conclude that (i) the new improved CCA scheme is an effective prediction tool, (ii) the error estimate of the new prediction tool is crucial at finding the optimal ensemble forecasting, (iii) other predictors, such as sea level pressure and soil moisture, should be considered, (iv) the GCM simulation data may be used to find the CCA statistical structure, and (v) further research is needed to determine the optimal EOF truncation order.

Of course, precipitation over some areas are not directly related to SST anomalies, but more related to SLP or other predictors. Zorita et al. (1992) also used CCA to study the precipitation predictability of Iberian precipitation from both SLP and SST data. They found that the dominant Iberian precipitation variability is caused by the westerly wind and

hence SLP patterns, but not the regional or remote SST anomalies. The same can be true for some regions of the US precipitation. However, our new CCA model has much flexibility and can easily include several predictors. Using multiple predictor can be an effective way to increase the forecasting skill. Therefore, a careful tuning of the model and a careful selection of predictors, such as soil moisture and SLP, can enhance the skill of prediction. Thus, our next project is to include other predictors in our optimal ensemble forecasting model.

The prediction skill can be further enhanced by using an extrapolation type of optimal combination of forecasts. Some of the forecasts with negative temporal correlation signs can also help with enhancing the skill of the optimal ensemble forecasting skill. Some of the weights of the forecasts can be negative. The weights are computed from a linear regression scheme similar to that of Krishnamurti et al. (2000). The assumption of independence of the forecasting errors of different models is no longer needed. This work is deferred to future studies.

Acknowledgments. The authors would like to acknowledge the Global Modelling and Data Analysis Program and the Tropical Rainfall Measuring Mission of the NASA Earth Science Enterprise for partial support of this study. This work was performed while Professor Shen held a National Research Council-(GSFC/NASA) Research Associateship. This work was partially supported by a subvention grant from the Atmospheric Environment Service of Environment Canada and a research grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Barnett, T.P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825-1850.

Barnston, A.G., and T.M. Smith, 1996: Specification and prediction of global surface temperature and precipitation from global SST using CCA. *J. Clim.*, **9**, 2660-2697.

Chen, M., P. Xie, J.E. Janowick and P.A. Arkin, 2001: Global land precipitation: 1 50-year monthly analysis based on gauge observations. *J. Hydrometeorology*, submitted for publication.

Higgins, W., J.E. Janowiak, and Y.-P, Yao, 1996: A gridded hourly precipitation data base for the United States (1963-1993), NCEP/Climate Prediction Center ATLAS No. 1, Camp Springs, MD 20746.

Kim, K.-Y., and G.R. North, 1998: EOF-based linear prediction algorithm: Theory. *J. Clim.*, **11**, 3046-3056.

———, and ———, 1999: EOF-based linear prediction algorithm: Examples. *J. Clim.*, **12**, 2076-2092.

Krisnamurti, T.N., C.M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Willford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Clim.*, **23**, 4196-4216.

Lau, K.M., and H.-T. Wu, 1999: Assessment of the impacts of the 1997-1998 El Nino on the Asian-Australia monsoon. *Geophys. Res. Lett.*, **26**, 1747-1750.

North, G.R., T.L. Bell, R.F. Cahalan, and F.J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Monthly Wea. Rev.*, **110**, 699-706.

Ropelewski, C.F. and M.S. Halpert, 1986: North American precipitation and temperature patterns associated with El Nino/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352-2362.

Reynolds, R.W., and T.M. Smith, 1994: Improved global sea surface temperature analysis using optimum interpolation. *J. Clim.*, **7**, 929-948.

Shen, S.S., T.M. Smith, C.F. Ropelewski and R.E. Livezey, 1998: An optimal regional averaging method with error estimates and a test using tropical Pacific SST Data. *J. Climate*, **11**, 2340-2350.

Smith T.M., R.W. Reynolds, R.E. Livezey, D.C. Stokes, 1996: Reconstruction of historical sea surface temperature using empirical orthogonal functions. *J. Climate*, **9**, 1403-1420.

van den Dool, H.M., J. Hopinggarner, E. O'Lenic, A.J. Wagner, J. Huang, and R. Churchill, 1997: Second annual review of skill of CPC real time long lead prediction, *Proceedings of the Twenty-Second Annual Climate Diagnostics and Prediction Workshop*, Berkeley, California, pp.10-14.

von Storch, H., and F.W. Zwiers, *Statistical Analysis in Climate Research*, Cambridge University Press, 1999, Chs. 14 and 15.

Xue, Y., A. Leetmaa, and M. Ji, 2000: ENSO prediction with Markov models: the impact of sea level. *J. Climate*, **13**, 849-871.

APPENDIX A CONVENTIONAL CCA

This appendix describes the CCA scheme without consideration of the area-factor. This is the scheme widely used in social sciences, finance, climatology, statistics, and other fields. Emphasized here is the symmetry of the matrix, whose eigenvalue is the maximal correlation between two fields. This symmetry guarantees that the eigenvalues are real numbers. However, in many meteorological literature, this symmetry has been ignored.

A.1. Conventional CCA and an example

The random vectors $\mathbf{X}$ and $\mathbf{Y}$ represent climate parameters at different points. Linear combinations of the elements of $\mathbf{X}$ and those of $\mathbf{Y}$ form two scalar random variables. The linear combinations can have various kinds of physical interpretations, depending on how a linear combination is made. For example, the following linear combination from uniform weights

$$\bar{X} = \sum_{i=1}^{N} \frac{1}{N} X_i$$

is the arithmetic average of the field and can be simply written as $\mathbf{a}'\mathbf{X}$, where

$$\mathbf{a}' = (1/N \;\; 1/N \;\; \cdots \;\; 1/N) \;\text{ (a row-vecter)}, \quad \mathbf{X} = (X_1 \;\; X_2 \;\; \cdots \;\; X_N)' \;\text{ (a column-vecter)}.$$

The objective of CCA is to find vectors $\mathbf{a}$ and $\mathbf{b}$ such that

$$\rho = \mathrm{Corr}(U, V) \tag{95}$$

be maximized, where

$$U = \mathbf{a}'\mathbf{X}, \quad V = \mathbf{b}'\mathbf{Y} \tag{96}$$

are called canonical variables.

If the correlation is large, say, 0.7, we may say that the parameters $\mathbf{X}$ and $\mathbf{Y}$ have high mutual predictability. Then, a multiple-variate regression procedure may be used for prediction.

The CCA scheme is derived from

$$\max \mathrm{Cov}(U, V), \tag{97}$$

$$\text{under the condition:} \qquad \mathrm{Var}(U) = \mathrm{Var}(V) = 1. \tag{98}$$

To proceed with the computing, the following covariance matrices are either given or computed from data:

$$\Sigma_{XX} = \mathrm{Cov}(\mathbf{X}, \mathbf{X}), \quad \Sigma_{YY} = \mathrm{Cov}(\mathbf{Y}, \mathbf{Y}), \quad \Sigma_{XY} = \mathrm{Cov}(\mathbf{X}, \mathbf{Y}), \quad \Sigma_{YX} = \Sigma'_{XY}. \tag{99}$$

The conventional CCA scheme requires that Σ_{XX} and Σ_{YY} be of full rank and can be diagonalized. However, because of the temporal length of the climatological data streams is often shorter than the spatial data points, the covariance matrices are often not full rank. Hence, the spectral method is used to overcome this problem, since the eigenvalue problem of the CCA in the EOF space is a problem of symmetric matrix. As a matter of fact, the CCA problem in the EOF space is symbolically the same as the SVD problem in the physical space.

The CCA algorithm includes the following three steps.

Step 1. Diagonalize matrices $\Sigma_{XX} = P'_X \Lambda_X P_X$ and $\Sigma_{YY} = P'_Y \Lambda_Y P_Y$, and find

$$\Sigma_{XX}^{-1/2}, \Sigma_{YY}^{-1/2}, \Sigma_{XX}^{1/2}, \Sigma_{YY}^{1/2},$$

where Σ_{XX}^m is defined as $P'_X \Lambda_X^m P_X$ for a real number m, and Σ_{YY}^m is similarly defined.

Step 2. Solve the eigenvalue problem

$$\Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1/2} \mathbf{e} = \rho^2 \mathbf{e}. \tag{100}$$

Step 3. Find the canonical variables. The linear combination vectors are

$$\mathbf{a} = \Sigma_{XX}^{-1/2} \mathbf{e}, \tag{101}$$

$$\mathbf{b} = \frac{1}{\rho} \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1/2} \mathbf{a}. \tag{102}$$

The canonical variables are

$$U = \mathbf{a}'\mathbf{X}, \qquad V = \mathbf{b}'\mathbf{Y}, \tag{103}$$

and their variances are equal to one. Their maximal correlation is ρ_1, the first eigenvalue.

Example. Find the maximal correlation between the linear combinations of two two-dimensional vectors

$$\mathbf{X} = (X_1, X_2), \quad \mathbf{Y} = (Y_1, Y_2),$$

if the covariance matrices are as follows:

$$\Sigma_{XX} = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}, \tag{104}$$

$$\Sigma_{YY} = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}, \tag{105}$$

$$\Sigma_{XY} = \begin{bmatrix} 0.6 & 0.4 \\ 0.7 & 0.5 \end{bmatrix}. \tag{106}$$

The first step is to find eigenvalues and eigenvectors of Σ_{XX} and Σ_{YY}, so that the needed matrices can be prepared. We have the following

$$\Sigma_{XX}^{-1/2} = \begin{bmatrix} 1.1154 & -0.2989 \\ -0.2989 & 1.1154 \end{bmatrix}, \tag{107}$$

$$\Sigma_{YY}^{-1} = \begin{bmatrix} 1.0989 & -0.3297 \\ -0.3297 & 1.0989 \end{bmatrix}, \tag{108}$$

$$\Sigma_{XX}^{1/2} = \begin{bmatrix} 0.9659 & 0.2588 \\ 0.2588 & 0.9659 \end{bmatrix}. \tag{109}$$

The second step is to find the CCA matrix and solve the CCA eigenvalue problem.

$$\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1/2} = \begin{bmatrix} 0.2393 & 0.3216 \\ 0.3216 & 0.4347 \end{bmatrix} \tag{110}$$

The eigenvalues and eigenvectors of this matrix are

$$\lambda_1 = \rho_1^2 = 0.6731, \quad \lambda_2 = \rho_2^2 = 0.0009, \tag{111}$$

and

$$[\mathbf{e}_1 \ \ \mathbf{e}_2] = \begin{bmatrix} 0.5955 & 0.8033 \\ 0.8033 & -0.5955 \end{bmatrix}. \tag{112}$$

The third step is to find the canonical variables. For the first eigen-mode, we have

$$[\mathbf{a}_1 \ \ \mathbf{a}_2] = \Sigma_{XX}^{-1/2}[\mathbf{e}_1 \ \ \mathbf{e}_2] = \begin{bmatrix} 0.4241 & 1.0740 \\ 0.7180 & -0.9043 \end{bmatrix}, \tag{113}$$

$$[\mathbf{b}_1 \;\; \mathbf{b}_2] = \frac{1}{\rho_1}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1/2}[\mathbf{a}_1 \;\; \mathbf{a}_2] = \begin{bmatrix} 0.8016 & 0.6755 \\ 0.4039 & -0.9674 \end{bmatrix}. \tag{114}$$

The canonical variables are

$$U_1 = \mathbf{a}_1'\mathbf{X} = 0.4241X_1 + 0.7180X_2, \quad U_2 = \mathbf{a}_2'\mathbf{X} = 1.0740X_1 - 0.9043X_2, \tag{115}$$

$$V_1 = \mathbf{b}_1'\mathbf{Y} = 0.8016Y_1 + 0.4039Y_2, \quad V_2 = \mathbf{b}_2'\mathbf{Y} = 0.6755Y_1 - 0.9674Y_2. \tag{116}$$

The maximal correlation between the canonical variables U_1 and V_1 is $\rho_1 = 0.8204$. This result can be verified by using the above two equations and the given covariance matrices at the beginning of the example:

$$\begin{aligned} & \mathrm{Cov}(U_1, V_1) = \langle U_1 V_1 \rangle \\ = \;& 0.4241 \cdot 0.8016\langle X_1 Y_1 \rangle + 0.4241 \cdot 0.4039\langle X_1 Y_2 \rangle + \\ & 0.0.7180 \cdot 0.0.8016\langle X_2 Y_1 \rangle + 0.7180 \cdot 0.4039\langle Y_2 Y_2 \rangle \\ = \;& 0.8204. \end{aligned}$$

This can also be verified in matrix notation:

$$[\mathbf{a}_1 \;\; \mathbf{a}_2]'\Sigma_{XY}[\mathbf{b}_1 \;\; \mathbf{b}_2] = \begin{bmatrix} \langle U_1 V_1 \rangle & \langle U_1 V_2 \rangle \\ \langle U_2 V_1 \rangle & \langle U_2 V_2 \rangle \end{bmatrix} = \begin{bmatrix} 0.8204 & 0.0000 \\ 0.0000 & 0.0295 \end{bmatrix}. \tag{117}$$

It is also interesting to see how the random variable $\mathbf{X}$ are represented by the canonical variables. To do so, one has to find all the canonical variables, i.e., U_1 and U_2 in the present case. The formula is

$$\mathbf{X} = \Sigma_{XX}^{1/2}(\mathbf{e}_1 \; \mathbf{e}_2)\mathbf{U}, \tag{118}$$

where $\mathbf{U} = (U_1 \; U_2)'$ is the vector of canonical variables. The above can be written as

$$\mathbf{X} = U_1\mathbf{E}_1 + U_2\mathbf{E}_2, \tag{119}$$

where

$$\mathbf{E}_1 = \Sigma_{XX}^{1/2}\mathbf{e}_1 = (0.7831 \;\; 0.6218)', \quad \mathbf{E}_2 = \Sigma_{XX}^{1/2}\mathbf{e}_2 = (0.9301 \;\; -0.3673)' \tag{120}$$

are called canonical patterns, as the physical field $\mathbf{X}$ is represented by the product between the patterns and the canonical variables. These two vectors are usually not orthogonal.

Similarly, the random variable $\mathbf{Y}$ can be represented by the canonical variables and patterns:

$$\mathbf{Y} = V_1\mathbf{F}_1 + V_2\mathbf{F}_2, \tag{121}$$

where

$$\mathbf{F}_1 = (0.9228 \quad 0.3853)', \quad \mathbf{F}_2 = (0.6444 \quad -0.7647)'. \tag{122}$$

APPENDIX B ANOTHER CCA SCHEME OF TWO CONTINUOUS FIELDS

This CCA algorithm retains the covariance functions in the physical space and is different from the one explained in Sections 2 and 3, which is in the spectral space. When a climate field is not expanded into spectral spaces, the current method may be a useful alternative to the CCA algorithm described in Section 2.

The weight functions are expanded in terms of EOFs:

$$u(\mathbf{x}) = \sum_n u_n \psi_n(\mathbf{x}) / \sqrt{\lambda_n^X}, \tag{123}$$

$$v(\mathbf{y}) = \sum_m v_m \phi_m(\mathbf{y}) / \sqrt{\lambda_m^X}. \tag{124}$$

The eigenfunctions are then determined by the coefficients u_n and v_m, which are, in turn, determined by linear algebraic equations.

With the above expansions, the equations of the eigenvalue problems (10) and (11) become

$$\int_{\Omega_Y} d\Omega_Y \; \Sigma_{XY}(\mathbf{x}, \mathbf{y}) \sum_n v_n \phi_n(\mathbf{y}) / \sqrt{\lambda_n^Y} + 2\zeta \sum_m u_m \sqrt{\lambda_m^X} \psi_m(\mathbf{x}) = 0, \tag{125}$$

$$\int_{\Omega_X} d\Omega_X \; \Sigma_{XY}(\mathbf{x}, \mathbf{y}) \sum_m u_m \psi_m(\mathbf{x}) / \sqrt{\lambda_m^X} + 2\eta \sum_l v_l \sqrt{\lambda_l^Y} \phi_l(\mathbf{y}) = 0. \tag{126}$$

Using the orthonormal condition of the EOFs, one can derive the following from equation (125) above:

$$u_m = -\frac{1}{2\zeta} \sum_n C_{mn} v_n, \tag{127}$$

where

$$C_{mn} = \int_{\Omega_X} d\Omega_X \int_{\Omega_Y} d\Omega_Y \; \Sigma_{XY}(\mathbf{x}, \mathbf{y}) \frac{\psi_m(\mathbf{x})}{\lambda_m^X} \frac{\phi_n(\mathbf{y})}{\lambda_n^Y}. \tag{128}$$

Substituting these two formulas into (126), one obtains the eigenvalue problem for the coefficient-unit vector $(v_1, v_2, \cdots)$,

$$\sum_{n=1}^{\infty} M_{ln} v_n = \lambda \, v_l. \tag{129}$$

where

$$\lambda = 4\zeta\eta, \tag{130}$$

$$M_{ln} = \sum_{m=1}^{\infty} C_{ml} C_{mn}. \tag{131}$$

The coefficient matrix $[M_{ln}]$ is symmetric, hence the eigenvalues are real and the eigenvectors are orthogonal. Similarly, an eigenvalue problem for the unit vector $(u_1, u_2, \cdots)$ can be obtained.

The eigenvalue λ is directly related to the maximal correlation (ρ) between U and V. As a matter of fact,

$$\lambda = \rho^2. \tag{132}$$

The proof is given below.

The correlation is

$$\begin{aligned} \rho &= \mathrm{Corr}(U, V) \\ &= \int_{\Omega_X} d\Omega_X \int_{\Omega_Y} d\Omega_Y \; \Sigma_{XY}(\mathbf{x}, \mathbf{y}) u(\mathbf{x}) v(\mathbf{y}) \\ &= \sum_{m,n} C_{mn} u_m v_n. \end{aligned} \tag{133}$$

Formula (127) leads to

$$\sum_m u_m^2 = -\frac{1}{2\zeta} \sum_m u_m \sum_n C_{mn} v_n, \tag{134}$$

or

$$1 = -\frac{1}{2\zeta} \rho, \tag{135}$$

since $(u_1, u_2, \cdots)$ is a unit vector.

Similarly, one can derive

$$1 = -\frac{1}{2\eta} \rho, \tag{136}$$

The product of the formulas (135) and (136) yields

$$\rho^2 = 4\zeta\eta = \lambda. \tag{137}$$

The formulas (135) and (136) also imply that the two Lagrange multipliers are equal to each other,

$$\eta = \zeta = -\frac{\rho}{2}. \tag{138}$$

This gives the physical meaning of the Lagrange multipliers, which is related to the correlation between the two fields under consideration.

APPENDIX C OPTIMAL EOF-TRUNCATION ORDER

Optimal truncation orders p and q should take values such that the observed data best support the physics. An expansion of a higher order contains too much noise, and that of a

lower order leaves out some signal. From computational aspect, the first H CCA eigenvalues $\rho_h(h = 1, \cdots, H)$ are equal to one, where $H = p + q - K + 1$ and K is the temporal length of the data streams. According to the CCA prediction error, the forecasting of the first h modes have zero error, which is, of course, impossible. Thus, p and q cannot be too large. If $p = q$, then the upper bound is $p < (K-1)/2$. If there are 20 years of data, then $p < 10$. To give some insight about the truncation, we include some further analysis below.

When EOFs form a complete basis for a linear space, then the corresponding field can be exactly expressed by an infinite EOF series

$$X(x,t) = \sum_{n=1}^{\infty} \sqrt{\lambda_n^X} X_n(t)\psi_n(x). \tag{139}$$

Of course, this is for a continuous field. The completeness condition of the EOF basis is expressed by

$$\sum_{n=1}^{\infty} \psi_n(x)\psi_n(x') = \delta(x - x'). \tag{140}$$

With a discrete field, the climate is contained in a finite dimensional space. Suppose that there are J_X discrete points. Then the field X must be contained in a J_X-dimensional vector space, denoted by $\mathcal{J}$. The EOF expression of the field should contain as little noise as possible. With true EOFs, the climate signal may be contained in a sub-space spanned by N_c dimensional sub-space, denoted by $\mathcal{S}_c$, where $N_c < J_X$. Hence, $\mathcal{J} = \mathcal{S}_c + \mathcal{S}_e$, where $\mathcal{S}_e$ contains no climate signal. The EOFs computed from data are the approximate EOFs. The first a few computed EOFs may agree with the true EOFs and contain climate signals. The higher order EOFs, whose eigenvalues are close to each other, are far away from the true EOFs, according to the North's rule-of-thumb (North et al., 1982). Therefore, the sub-space spanned by the computed EOFs intersects partly with $\mathcal{S}_c$ and partly with $\mathcal{S}_e$. The climate signal sub-space is denoted by $\mathcal{S}_s$ (sample signal represented by the first p-EOFs) and $\mathcal{S}_n$ (sample noise represented by higher order EOFs). Only the EOFs in the signal sub-space are used in the forecasting analysis. Some properties of $\mathcal{S}_s$ are described below.

Observed data value at only the station points or grid points. If the EOFs can form a complete basis in a vector space J_X, which is the total number of grid points for parameter X, then the field can still be expressed by the EOFs:

$$X(i,t) = \sum_{n=1}^{J_X} \sqrt{\lambda_n^X} X_n(t)e_n(i). \tag{141}$$

Here, the covariance matrix is defined by

$$\langle X(i,t)X(j,t)\rangle = \frac{1}{K}\sum_{t=1}^{K} X(i,t)X(j,t), \tag{142}$$

where K is the temporal length of the data stream. The vector $e_n(i)$ is the nth eigenvector of this covariance matrix, hence is the nth EOF. When $K \geq J_X$, the covariance matrix is of full rank, and the EOFs form a complete space. The complete condition is

$$\sum_{n=1}^{J_X} e_n(i)e_n(j) = \delta_{ij}. \tag{143}$$

It is also interesting to know that

$$\frac{1}{J_X}\sum_{n=1}^{J_X} X_n(s)X_n(t) = \delta_{st}. \tag{144}$$

and

$$\frac{1}{K}\sum_{t=1}^{K} X_m(t)X_n(t) = \delta_{mn}. \tag{145}$$

When $K < J_X$, the rank of the covariance matrix is at most K. The covariance matrix has at most K non-zero eigenvalues, and hence K definite eigenvectors. These eigenvectors cannot form a complete basis in the J_X-dimensional vector space. Hence the EOF expansion of the field is equal to the field only in this K years, but they are not equal to each other in other years. The equality that holds in the K-data years can be proved from the calculation algorithm of the EOFs.

The EOFs are computed from

$$\sum_{j=1}^{J}\left(\frac{1}{K}\sum_{t=1}^{K} X(i,t)X(j,t)\right)e_n(j) = \lambda_n e_n(i). \tag{146}$$

The left hand side of the above may be written as

$$\sum_{t=1}^{K} X(i,t)\frac{1}{K}\sum_{j=1}^{J}(X(j,t)e_n(j)) = \sum_{t=1}^{K} X(i,t)\frac{1}{K}\sqrt{\lambda_n}X_n(t). \tag{147}$$

Combining the above two equations, one obtains

$$\sum_{t=1}^{K} X(i,t)\frac{1}{K}X_n(t) = \sqrt{\lambda_n}e_n(i). \tag{148}$$

The principal components satisfy that

$$\frac{1}{K}\sum_{n=1}^{K} X_n(s)X_n(t) = \delta_{st} \tag{149}$$

and

$$\frac{1}{K}\sum_{t=1}^{K} X_m(t)X_n(t) = \delta_{mn}. \tag{150}$$

Thus, the inverse matrix of

$$\frac{1}{K}\left[X_n(t)\right]$$

is

$$\left[X_n(t)\right]'.$$

With this inversion, eq. (148) becomes

$$X(i,t) = \sum_{n=1}^{K} \sqrt{\lambda_n^X} X_n(t) e_n(i). \tag{151}$$

According to North's rule-of-thumb, when two sample-eigenvalues are close to each other, the sampling error for EOFs is very large and the mixture of two or more true EOFs into a sample EOF occurs. The EOFs are defined as the best approximation of the field by a linear combination of EOFs. The best approximation is measured by the ensemble average. That is,

$$\langle \sum_{j=1}^{J} \left(X(j,t) - \sum_{n=1}^{p} \sqrt{\lambda_n} X_n(t)\psi_n(j) \right)^2 \rangle. \tag{152}$$

When this is approximated by K years of data, we have the following problems:

(a) The approximation most likely has a large error beyond the K years.

(b) The mixture of two or more modes destroys the best approximation. Suppose that the sample EOF $e_4(j)$ is a mixture of $\psi_4(j)$ and $\psi_5(j)$. Then, there exist two constants a_4 and a_5 such that

$$e_4(j) = a_4\psi_4(j) + a_5\psi_5(j).$$

The projection of the field onto e_4 becomes the projections onto ψ_4 and ψ_5. Because of a_4 and a_5, the ψ_4 and ψ_5 projections are not orthogonal and hence not the best approximation. North's rule-of-thumb indicates that the first a few sample EOFs do not likely have mixture and hence should be used as a good approximation to true EOFs. This is equivalent to say that one takes only the EOFs whose eigenvalues are sufficiently separated. These are usually the first a few sample EOFs. The higher order EOFs contain the smaller scales which require more than K years of data to resolve. This is a type of aliasing phenomenon. Large sampling size resolves smaller spatial scales.

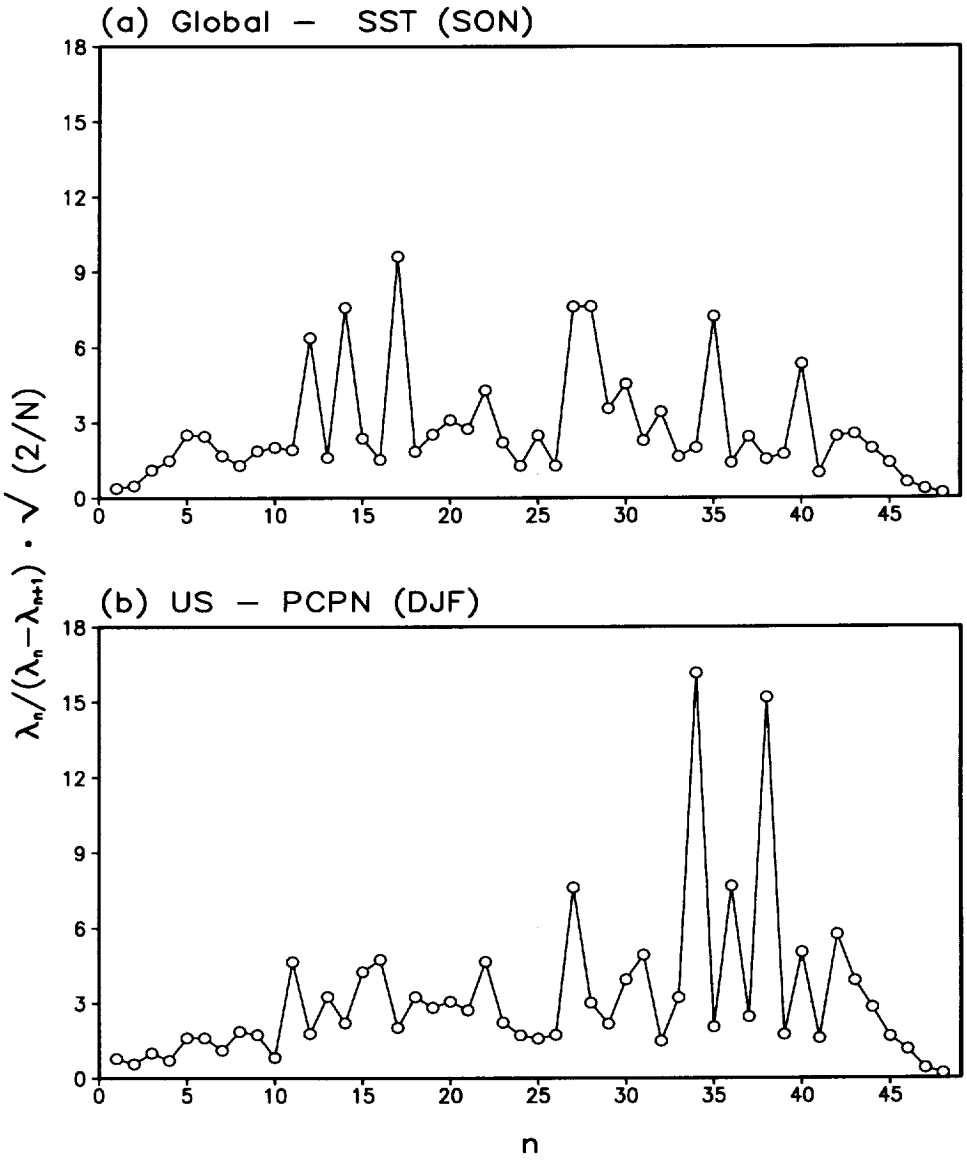

Figure 11: Separation of eigenvalues: (a) SON global SST, and (b) DJF US precipitation. The first 10 EOFs of the US precipitation are used in the forecasting experiments.

Unfortunately, there is no definite rule to determine the optimal truncation order p. The C-rule is used in this paper. For a given constant C, if

$$\frac{\lambda_n}{\lambda_n - \lambda_{n+1}}\sqrt{2/K} < C, \tag{153}$$

then the nth EOF is included. The optimal value p is equal to the maximal n. In this paper we choose $C = 3$ for both SST and precipitation. Examples of the separation of egenvalues for SST and US precipitation are displayed in Fig. 11. Thus, the SON global SST may use 11

modes. However, there is a strong evidence of mode mixture for the SON global Pacific SST between the $12th$ and $13th$ modes. Thus it is preferable to discard the modes whose order is higher than 12. The DJF precipitation's mode mixture happens at mode 11. So, modes 1-10 may be used for forecasting. However, our CCA algorithm requires the same number of modes for both predictor and predictand. Hence we have chosen 10 modes from both SST and precipitation EOFs.